DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

The Discovery and Characterization of NAD-Linked RNA

Citation

Chen, Ye Grace. 2012. The Discovery and Characterization of NAD-Linked RNA. Doctoral dissertation, Harvard University.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:9550687

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

© 2012 by Ye Grace Chen
All rights reserved

Dissertation Advisor: Professor David R. Liu Ye Grace Chen

The Discovery and Characterization of NAD-Linked RNA

ABSTRACT

Over the past few decades, RNA has emerged as much more than just an intermediary in biology's central dogma. RNA is now known to play a variety of catalytic, regulatory and defensive roles in living systems as demonstrated through the discoveries of ribozymes, riboswitches, microRNAs, small interfering RNAs, Piwi-interacting RNAs, small nuclear RNAs, clusters of regularly interspaced short palindromic repeat RNAs and long non-coding RNAs. In contrast to the functional diversity of RNA, the chemical diversity has remained primarily limited to canonical polyribonucleotides, the 5' cap on mRNAs in eukaryotes, modified nucleotides and 3'-aminoacylated tRNAs. This disparity coupled with the powerful functional properties of small molecule-nucleic acid conjugates led us to speculate that novel small molecule-RNA conjugates existed in modern cells, either as evolutionary fossils or as RNAs whose functions are enabled by the small molecule moieties.

We developed and applied a nuclease-based screen coupled with high-resolution liquid chromatography/mass spectrometry analysis to detect novel small molecule-RNA conjugates, broadly and sensitively. We discovered NAD-linked RNA in two types of bacteria and further characterized the small molecule and RNA in *Escherichia coli*. The NAD modification is found on the 5' end of RNAs between 30 and 120 nucleotides long, and is surprisingly abundant at around 3,000 copies per cell. Subsequent experiments to characterize further NAD-linked RNA have been undertaken, including sequencing the RNAs to which NAD is attached and elucidating the biological functions of the small molecule-RNA conjugate.

The development and application of a screen to detect novel nucleotide modifications that is independent of structure or biological context has the potential to increase our understanding of the functional and chemical diversity of RNA. The discovery and biological characterization of NAD-linked RNA can provide new examples of RNA biology and offer insight into the RNA world.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Graduate school has been a journey, and the people I've met along the way and spent time with have been an instrumental part of my experience here. There is a large community of people who have helped me to get to this point, and I am deeply grateful for all of the ways that they've contributed to my academic and personal lives.

The Liu lab is an amazingly resourceful place to work, both intellectually and instrumentally. I thank David for providing us this environment to conduct research and pursue interesting questions. David has been an encouraging and supportive advisor, patient when things were taking longer than expected and optimistic when things weren't going smoothly. His enthusiasm and positivity about this project has carried us farther than we would have gone otherwise.

The NAD-linked RNA project has received a lot of support from fellow lab members. I thank the past and present members of TEAM SMRC: Walter Kowtoniuk, Yinghua Shen, Aaron Leconte, Diane Truong and Christoph Dumelin, for all of the brainstorming, discussion and instrument troubleshooting. Courtney Yuen and Yevgeny Brudno have offered helpful suggestions about the project from the beginning. Vikram Pattanayak and Jacob Carlson have always taken the time to go over my data and provide interpretation and analysis. Thanks to effective molarity, Lynn McGregor and Kevin Davis have made mundane pipetting entertaining and I am privileged to be part of their daily lives. I am appreciative for the friendship of all of these people and I thank the entire Liu lab for the fun times.

My committee members have been wonderful; I thank Chris Walsh, Alan Saghatelian and Jon Clardy. Despite their busy schedules, they have always made time to discuss project

details and future directions with me. Each time that we met, my committee has provided helpful advice, asked thoughtful questions and offered kind encouragement.

I have had been blessed by a wonderful group of friends in the Cambridge area. I was lucky enough to be matched with Dean Sheila Thomas and Christine Kiely for the HGWISE mentoring program. We have met regularly over the past few years to share stories and laughter. I thank Alex Speed and Josh Paulk for sharing their perspectives during all the lunches and conversations. My sincerest thanks goes to the people who have played outside with me, cooked for me, or both. Those times have sustained me and always been uplifting. Linda Liang, Theresa Liang, Audrey Kung, Sarah Mahoney and Susanna Vagt have been my ladies.
I thank A. John Lian, Alex Chan, Gary Sing, Tout Wang, Sam Wong, Alex Frenzel, Peter Blair, Jason Chen, and Michael Feng for all the care that they've extended in my direction.

My teachers in Davis, California and research mentors afterward were instrumental in helping me get to Harvard. I especially thank Mr. Larry Duque (Valley Oak Elementary School), Mrs. Moldenhauer (Davis High School), Professor Julie Leary (UC Davis), Professor Tom Alber (UC Berkeley), Kana Yamamoto (Bristol-Myers Squibb), Professor Bing Zhou (Tsinghua University), Kean Goh and Jonathan Sullivan (California Environmental Protection Agency). They remind me of the joy of learning and continue to teach me how to maintain a broad perspective.

I am so glad that Rou-Jia Sung and I ended up in labs next to each other here. Thank you for reminiscing about the California sunshine with me over the last six years and for seeing Harvard through Cal eyes too. I thank Lindsay White for all of the hostessing. Sometimes your kindness is so stunning that I am humbled. Anais Lim has been my Mujer since freshmen year at Berkeley. Even though we've lived our lives almost oppositely on opposite coasts, you

understand me. Thank you for joining me for boba and mangoes even when you had better things to do.

In terms of my family, words are not enough to express my gratitude for their love and care, 捧, of me. "How does a blade of grass thank the sun?" My parents have supported and encouraged me through the roughest and best times. Thank you for being my fiercest advocates. Jane is the best younger sister and older brother I could ask for, humoring my antics with a tolerance that far surpasses her years. Her vivacity and resilience have been an example to me of how to find joy in all situations. You are my superlative. True story. And finally, my grandparents were among the first to princess me. Thank you for understanding that my legs are meant for dancing and for carrying me all of those years.

for my grandparents, all four of them

捧：双手托着， 小心地保护着

pěng: carry in both hands, hold closely and preciously

Chapter One:

The Functional and Chemical Diversity of Cellular RNA

1.1 The Functional Diversity of Cellular RNA

The functional roles of RNA have expanded tremendously over the past few decades to encompass much more than just the central dogma. Ribozymes, riboswitches, microRNAs, small interfering RNAs, transcriptional regulators, and long non-coding RNAs are all examples of RNAs that are thought to play a broad range of catalytic, sensing, regulatory, or defensive roles in the cell. These recently discovered novel RNAs have illuminated new mechanisms of important biological processes.

1.1.1 RNA as Key Component of the Cellular Protein Translation Machinery

The central dogma describes RNA's cellular role as transferring genetic information from DNA into proteins. Indeed, the first few discoveries of functional RNA concluded that it was involved with protein synthesis. The ribosome was found to contain large RNAs, and abundant tRNA was necessary for cellular production of proteins. mRNA provided the connection between DNA and proteins by acting as a substrate for the ribosome to template translation.

By 1941, scientists learned that proteins were being synthesized in the presence of abundant RNAs. Another two decades passed before the first functional role for RNA was discovered as an essential component to the translation machinery. George Emil Palade observed a "small particulate component of the cytoplasm" in guinea pig liver cells using the electron microscope. He saw molecules that were either free in the cytoplasm or bound to the cisternal membranes of the rough portion of the endoplasmic reticulum (ER). Palade then moved to guinea pig pancreatic cells because they are very efficient protein producers and are packed with endoplasmic reticulum cirsternae studded with "microsomes" – what the "small particulates" were called.

Together with Philip Siekevitz, Palade isolated the pancreas from guinea pigs and fractionated the cellular components using a sucrose gradient. They examined the different fractions using electron microscopy and established that microsomes arise by a peculiar fragmentation of the ER during which membrane vesicles with their attached particles pinch off from ER cisterna without leakage of the cisternal content.[1] They used sodium deoxycholate to solubilize microsomal membranes, which allowed for the recovery of the attached particles by high-speed centrifugation. The particles were shown to be rich in protein and RNA, and hence, were named RNPs for ribonucleoprotein particles. Using biochemistry, cell fractionation procedures and electron microscopy, Palade and Siekevitz elucidated the function of the ribonucleoprotein particles as the exclusive site of protein synthesis and found that RNA is essential to the production of protein in the ribosome.[1-5]

During this time, there was additional evidence accumulating that implicated a role for cellular RNA in protein synthesis. The intermediate stages between amino acid activation and final incorporation into protein in the rat liver *in vitro* system offered unexplored regions in which to seek more direct evidence for a chemical association of RNA and amino acids. By 1953, Paul Zamecnik had succeeded in making the first cell-free system capable of carrying out net peptide bond formation through incorporation of ^{14}C-amino acids.[6] Using this system, he, Elizabeth Keller, and Mahlon Hoagland noticed that the RNA in a particular cytoplasmic fraction became labeled with ^{14}C-amino acids and that the labeled RNA was subsequently able to transfer the amino acids to microsomal protein in the presence of GTP.[7] From this, they concluded that the RNA, later named transfer RNA or tRNA, functions as an intermediate carrier of amino acids in protein synthesis.

However, a missing connection between the information carrier (DNA) and protein synthesis remained. While there was still no direct evidence, scientists hypothesized that the order of nucleotides in DNA determined the order of amino acids in a protein. Francis Crick proposed that RNA may act as the carrier of genetic information to direct protein synthesis.[8] He predicted that an "adaptor" molecule that mediates between the genetic information in DNA and the functional amino acids of proteins.

Experimental evidence for mRNA came in 1961 when researchers cultured *E. coli* on media containing heavy carbon and nitrogen isotopes resulting in *E. coli* with labeled ribosomes.[9] Then, the cells were infected with T2 bacteriophage and transferred to media containing $^{32}P0_4$ as the sole phosphorous source. All of the RNA that was synthesized following the infection contained the ^{32}P label and associated to the heavy ribosomes upon density gradient centrifugation. During infection, there is vigorous synthesis of phage protein and DNA, but no change in bacterial RNA content. Therefore, researchers concluded that there was an unstable RNA intermediate that was not rRNA or tRNA that allowed for protein synthesis. This discovery revealed a fundamental mechanism for gene action: the coding sequences of genes are copied into short-lived RNAs that are transported out of the nucleus into the cytoplasm, where they are translated into proteins. Because such RNAs carry information from genes in the nucleus to the cytoplasm, they are designated as messenger RNAs.

During protein synthesis, tRNAs deliver amino acids to the ribosome, where the ribonucleoprotein complex links amino acids together to form proteins. mRNA provides the template for tRNA to base pair with in order to translate nucleic acids into amino acids. tRNA, rRNA and mRNA form the foundation of biology's central dogma where genetic information from DNA is converted into functional proteins.

1.1.2 RNA as Key Component of the Complex for Protein Membrane Integration

Signal recognition particle (SRP) RNA was first detected in avian and murine oncogenic RNA virus particles.[10] Researchers purified the virus and extracted total RNA. After chromatography of the nucleic acids, they found a 7S RNA that had not previously been described. Subsequently, SRP RNA was found to be a stable component of uninfected HeLa cells where it associated with membrane and polysome fractions.[11,12] In 1980, cell biologists purified from canine pancreas an 11S "signal recognition protein" (fortuitously also abbreviated "SRP") which promoted the translocation of secretory proteins across the membrane of the ER.[13] It was then discovered that SRP contained an RNA component.[14]

SRP is an abundant, cytosolic, universally conserved ribonucleoprotein that recognizes and targets specific proteins to the endoplasmic reticulum in eukaryotes and the plasma membrane in prokaryotes. The function of SRP was discovered by the study of processed and unprocessed immunoglobulin light chains,[15] which were produced in cell-free and reconstituted systems. After introducing radioactive amino acids into the biological systems during protein synthesis, researchers observed two different molecular weight products. Digestion of the products showed that the N-terminus was altered, so they proposed that the light chains are initially synthesized as a larger precursor of slightly higher molecular weight and subsequently converted into the authentic product.

The SRP binds the N-terminal hydrophobic signal sequences in newly synthesized proteins in eukaryotes when they emerge from the ribosome. This interaction leads to the slowing of protein synthesis known as "elongation arrest," a conserved function of SRP that facilitates the coupling of the protein translation and the protein translocation processes. SRP then targets this entire complex (the ribosome-nascent chain complex) to the protein-conducting

channel in the ER membrane, which occurs via the interaction and docking of SRP with its cognate SRP receptor that is located in close proximity to the translocon.

1.1.3 RNA as Key Component of Ribosome Rescue

Transfer-messenger RNA (tmRNA, also known as 10Sa RNA and SsrA) is a bacterial RNA molecule with dual tRNA-like and mRNA-like properties.[16] The tmRNA forms a ribonucleoprotein complex (tmRNP) together with Small Protein B (SmpB), Elongation Factor Tu (EF-Tu), and ribosomal protein S1. In *trans*-translation, tmRNA and its associated proteins bind to bacterial ribosomes that have stalled in the middle of protein biosynthesis. The tmRNA recycles the stalled ribosome, adds a proteolysis-inducing tag to the unfinished polypeptide, and facilitates the degradation of the aberrant mRNA. In the majority of bacteria these functions are carried out by standard one-piece tmRNA. In other bacterial species, a permuted *ssrA* gene produces a two-piece tmRNA in which two separate RNA chains are joined by base-pairing.

tmRNA was first designated 10Sa RNA after a mixed "10S" electrophoretic fraction of *E. coli* RNA was further resolved into tmRNA and the similarly-sized RNase P RNA (10Sb).[17] The presence of pseudouridine in the mixed 10S RNA hinted that tmRNA has modified bases found also in tRNA. The similarity at the 3' end of tmRNA to the T stem-loop of tRNA was first recognized upon sequencing *ssrA* from *M. tuberculosis*.[18] Subsequent sequence comparison revealed the full tRNA-like domain formed by the 5' and 3' ends of tmRNA, including the acceptor stem with elements like those in alanine tRNA that promote its aminoacylation by alanine-tRNA ligase.[19] It also revealed differences from tRNA: the anticodon arm is missing in tmRNA, and the D arm region is a loop without base pairs.

1.1.4 RNA as Regulator of Cellular Gene Expression

Gene regulation is the process where cells and viruses regulate the way that information in genes is turned into gene products, which are typically proteins. RNAs in all kingdoms of life have been discovered and characterized to participate in gene regulation. The number of noncoding RNA genes, ones that produce functional RNA rather than encode for protein translation, has greatly increased over the past few decades. The clusters of regularly interspaced short palindromic repeat (CRISPR) RNA is thought to confer a defensive mechanism to its cells and serve as a "genetic memory" of the parasites they have encountered. Long noncoding RNA has a variety of functions, including imprinting and dosage compensation. Riboswitches bind small molecule metabolites to regulate it's own transcription and translation. Small RNAs such as antisense RNA, microRNA, piwi-interacting RNA and small interfering RNA function as guide RNAs within the broad phenomenon known as RNA silencing. All of these discoveries demonstrate that RNAs are quite diverse in their regulatory roles.

RNA interference (RNAi) is the induction of sequence-specific gene silencing by double-stranded RNA (dsRNA) found exclusively in eukaryotes. In the laboratory, RNAi is a powerful tool that makes gene inactivation possible in organisms that were not amenable to genetic analysis before. In nature, RNAi may both play an important biological role in protecting the genome against instabilities caused by transposons and repetitive sequences and be an ancient antiviral response/protection mechanism in both animals and plants.[20,21] Recent genetic analyses provide evidence that RNAi may also have integral functions in the regulation of endogenous genes.[22]

The RNAi process of dsRNAi-directed mRNA cleavage begins when Dicer or a Dicer homolog cleaves the dsRNA to 21-23 nucleotide long fragments.[23-26] The resulting small

interfering (siRNA) fragments are bound by RNAi-specific enzymes and are incorporated into a distinct RNA-induced silencing complex (RISC) that targets mRNA for degradation. The complex cleaves the mRNA in the center of the region recognized by the siRNA.

In addition to siRNAs, microRNAs (miRNAs) are another type of small RNA that participates in RNAi. The endogenous ~23 nt RNAs play important gene-regulatory roles in animals and plants by pairing to the mRNAs of protein-coding genes to direct their posttranscriptional repression.[27] miRNAs were discovered when two small regulatory RNAs, *lin-4* and *let-7*, were found to control the timing of larval development in the worm *Caenorhabditis elegans*.[28,29] In 1993, researchers found that LIN-14 protein abundance was regulated by a short RNA product encoded by the lin-4 gene. A 61-nucleotide precursor from the lin-4 gene matured to a 22-nucleotide RNA that contained sequences partially complementary to multiple sequences in the 3' UTR of the *lin-14* mRNA. This complementarity was both necessary and sufficient to inhibit the translation of the *lin-14* mRNA into the LIN-14 protein. Only in 2000 was a second RNA characterized: let-7, which repressed *lin-41*, *lin-14*, *lin-28*, *lin-42*, and *daf-12* expression during developmental stage transitions in *C. elegans*. *let-7* was soon found to be conserved in many species, including mammals, indicating the existence of a wider phenomenon.[30,31] miRNAs have since been found in plants, green algae, viruses, and more deeply branching animals.

miRNAs show very different characteristics between plants and metazoans. In plants, repressions on transcriptional level usually require perfect or near perfect target match, while mismatched target can lead to gene silence on translational level.[32] In metazoans, on the other hand, miRNA complementarity typically encompasses the 5' bases 2-7 of the microRNA, and one miRNA can target many different sites on the same mRNA or on many different mRNAs.[27]

Another difference is the location of target sites on mRNAs. In metazoans, the miRNA target sites are in the 3'-untranslated region of the mRNA, whereas in plants, targets are more often in the coding region itself. While the detailed mechanisms are slightly different in metazoans and plants, miRNAs base pair with mRNAs in both to direct posttranscriptional repression.

Piwi-interacting RNA (piRNA) are distinct from miRNA in size (26–31 nt rather than 21–24 nt), lack of sequence conservation, and increased complexity. However, like other small RNAs, piRNAs are thought to be involved in gene silencing, specifically the silencing of transposons since the majority of piRNas are antisense to transposon sequences.[33] piRNAs form RNA-protein complexes that have been linked to both epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells, particularly those in spermatogenesis.[34] They are the largest class of small non-coding RNA molecules that is expressed in animal cells.[33]

Like the other small RNAs, antisense RNA act by sequence complementarity to target RNAs (sense RNAs), which are primarily mRNAs encoding proteins of important functions. These small, diffusible and highly structured RNAs are transcribed from a promoter located on the opposite strand of the same DNA molecule, and are, therefore, fully complementary to their target RNAs. However, over the past years, a number of antisense RNAs were detected that are encoded in trans, having only partial complementarity to their target RNA and having more than one target. Antisense RNA control functions in all three kingdoms of life, but the majority of examples are known from bacteria.[35]

In contrast with short RNAs, long non-coding RNA (long ncRNAs or lncRNA) are in general considered as non-protein coding transcripts longer than 200 nucleotides that also function to orchestrate genetic regulatory outputs. Genetic studies identified a few lncRNA

genes involved in imprinting and other cellular processes, where the human X-(inactive)-specific transcript RNA is a canonical example.[36] This 17-kb lncRNA has a key role in dosage compensation and X-chromosome inactivation. Long ncRNAs have also been shown to modulate the function of transcription factors by several different mechanisms, including functioning themselves as co-regulators, modifying transcription factor activity, or regulating the association and activity of co-regulators. However, many of the functional roles for lncRNAs still remain mostly elusive.

1.1.5 RNA as Key Component of the Bacterial Defense System

The clusters of regularly interspaced short palindromic repeat (CRISPR)-based defense system protects many bacteria and archaea against invading conjugative plasmids, transposable elements, and viruses.[37] The cells acquire resistance by incorporating short stretches of invading DNA sequences in genomic CRISPR loci. These integrated sequences, ranging in size from 24 to 48 base pairs, are thought to function as a genetic memory that prevents the host from being infected by viruses containing this recognition sequence. A number of CRISPR-associated (*cas*) genes has been reported to be essential for the phage-resistant phenotype.[37] However, the molecular mechanism of this adaptive and inheritable defense system in prokaryotes has remained unknown.

1.1.6 RNA as Cellular Metabolite Sensor

Traditionally, proteins were the only macromolecules that had the ability to monitor the metabolic status of the cell. However, in 2002, researchers presented the first comprehensive proofs of multiple riboswitch classes, including protein-free binding assays that established

metabolite-binding riboswitches as a new mechanism of gene regulation.[38-41] Riboswitches reside in the leader sequences of numerous bacterial operons and control both transcription and translation by adopting alternative RNA structures, which can induce or prevent the formation of intrinsic terminators or ribosome binding site sequesters. Thus, depending on the configuration of the leader transcript, the same riboswitch can either be a repressor or an activator of a cognate gene.[42] Therefore, mRNAs that contain a riboswitch regulates its own activity levels in response to the concentration of its target molecule.

Prior to the discovery of riboswitches, accumulating evidence suggested that the mRNA might be involved in binding metabolites directly to effect their own regulation. There were conserved RNA secondary structures often found in the UTRs of the relevant genes and the success of procedures to create artificial aptamers.[43-46] Structurally diverse metabolites have since been discovered to bind to the leader sequences of numerous metabolic genes in both Gram-positive and Gram-negative bacteria.[42] Interestingly, riboswitches in Gram-negative bacteria tend to function via modulation of translation initiation, whereas in Gram-positive bacteria they predominantly function via transcription termination.[42]

The known riboswitches are highly conserved among bacteria, which argues for their ancient origin.[42] They may represent molecular fossils, a holdover from the RNA world. By contrast, since riboswitches are the fastest reacting regulatory systems because no intermediate factors are involved, various organisms could develop their own "modern" riboswitches for other processes associated with RNA.[42]

1.1.7 RNA as Biological Catalyst

Before the discovery of ribozymes, proteins were the only known biological catalysts. Researchers had hypothesized that RNA could act as a catalyst due to it's ability to form complex secondary structures.[47] In 1982, Tetrahymena was found to contain a self-splicing RNA without any associated protein.[48] The following year, catalytic activity was discovered in the RNA component of ribonuclease (RNase) P, providing the first example of a multiple-turnover enzyme using RNA-based catalysis.[49] Since then, other investigators have discovered additional examples of self-cleaving RNA or catalytic RNA molecules. The term ribozyme was developed for the general concept of an RNA molecule with enzyme-like activity.

RNA-based catalysts have similarities to their protein enzyme counterparts. Ribozymes can have substantial rate enhancements, comparing the rate constant for the self-cleaving hepatitis delta virus (HDV) of 10^2 - 10^4 s^{-1} with the maximal cleavage rate of RNase A of 1.4 x 10^3 s^{-1} at 25 °C.[50] Like proteins, ribozymes can also use cofactors such as imidazole during catalysis.[51,52] Furthermore, molecular structures have revealed that ribozymes and protein catalysts can fold into specific three-dimensional shapes that can have deep grooves and solvent-inaccessible active sites. These structures facilitate catalysis by orienting structures adjacent to catalytic groups and metal ions.[50]

The functional diversity of RNA has greatly increased over the past few decades. RNA has been found to participate in a variety of unexpected cellular roles such as catalysis, metabolite sensing, gene regulation and defense against foreign nucleic acids, in addition to its participation in protein synthesis. The versatility of RNA is surprising, especially given the limited chemical diversity of these nucleic acids.

1.2 The Chemical Diversity of Cellular RNA

In the six decades since their initial discovery,[53] researchers have identified nearly 100 modified nucleosides, spanning the three kingdoms of life almost all in RNA.[54-56] The location and distribution of the chemical modifications vary greatly between different RNA molecules, organelles and organisms. While modifying enzymes and biosynthetic pathways have been identified for some of the nucleosides, the biochemical and physiological functions of many of these modifications are still mostly unknown.

When modified RNAs were first discovered, researchers assumed that the RNA polymerase incorporated the modified nucleotide during transcription.[57] A few years later, researchers showed that methylation occurred on the polynucleotide level following primary transcript synthesis. In fact, all RNA modifications in *E. coli* with the exception of queuosine, are formed after transcription. The number of RNA modification enzymes is much higher than the number of RNA modifications present in *E. coli*.[55] One nucleoside modification often requires the action of several modification enzymes. Even proteins that are not directly interacting with RNA may be considered an RNA modification enzyme since they may synthesize a substrate that another RNA modification enzyme may require for activity.

Overall, RNA modifications may reinforce hydrogen bonds (such as pseudouridine in RNA double helices), improve base stacking (like m^5s^2U in the TΨC loop and hypermodified purines in the anticodon loop), alter the flexibility of the nucleoside (like 2'-O-methyl derivatives) or promote additional binding sites for metal ions (mostly Mg^{2+}).[57] As a result, fully modified RNA is more rigid and more resistant to thermal denaturation than unmodified RNA transcript. However, unmodified RNAs show little change in phenotype when compared with modified RNAs.

1.2.1 Chemical Modifications on tRNA

Of all the RNAs in the cell, tRNAs contain the highest frequency of known modifications. ~10 % of bacterial tRNA are modified nucleosides, whereas up to 25 % of the nucleosides in tRNA from eukaryotes are modified.[57] There are common modifications that belong to bacteria, eukaryotes and archaea, some of which are even present in similar positions. Therefore, unless there was convergent evolution, this pattern suggests that there may be a common origin for some of these modified nucleosides. These altered nucleosides may enhance tRNA's ability to interact with diverse molecules such as aminoacyl–tRNA ligases, elongation factors and ribosomal proteins as well as mRNA and rRNA.

In tRNA, the wobble position and 3' to the anticodon bases are most frequently modified, and there is a large variety of the modification at those two positions.[54] This suggests that their presence is important for the anticodon-codon interaction. There is evidence that 5-methyluridine and queosine derivatives prevent missense errors within a codon box.[57] The presence of some modified nucleosides is important for the recognition of aminoacyl-tRNAs so appropriate charging of a tRNA and translation efficient may be dependent on these modifications. The modified nucleosides may also increase translational fidelity. For example, 1-methylguanosine prevents tRNA from frameshifting, and lack of it reduces the growth rate considerably in yeast and is essential in *Streptococcus pneumoniae*.[57] The high frequency and diversity of chemical modifications in the tRNA anticodon loop may also affect tRNA structure to optimize interactions with mRNA, rRNA and translational proteins.

1.2.2 Chemical Modifications on rRNA

While the modifications in tRNA are chemically diverse, rRNA nucleosides are modified in only a few ways. The modifications are confined to isomerization of U to pseudouridine (Ψ), addition of H across the 5,6 double bond of U to form dihydrouridine (hU), and addition of methyl groups to the purine and pyrimidine rings and to the 2'-hydroxyl of ribose. Similar to tRNA, the distribution of modified nucleosides is not random in rRNA either.[58] Three-dimensional maps of the modified nucleotides in the ribosomes of *E. coli* and yeast reveal that most (~95 % in *E. coli* and 60 % in yeast) occur in functionally important and conserved regions. These include the peptidyl transferase center, the A, P and E sites of tRNA- and mRNA binding, the polypeptide exit tunnel, and interacting faces of subunits. For example, domain V in the large subunit has long been linked to peptidyl-transferase activity and tRNA binding, and is especially rich in modifications.[59] The correlations suggest that many ribosome functions benefit from nucleotide modification including having altered steric properties, different hydrogen-bonding potential, enhancement of local base stacking and increased structural rigidity for both single- and double-stranded regions.[58] Notably, modifications are essentially absent from areas dominated by ribosomal proteins: the external surfaces and periphery of the interface regions, which suggest that most RNA-protein interactions are not affected directly by modification.[60-62]

1.2.3 Chemical Modifications on mRNA

mRNAs in eukaryotes have a 5' cap that consists of a 7-methylguanylate (m^7G) that provides significant resistance to 5' exonucleases, facilitates transport from the nuclease, enhances translation efficiency and increases mRNA splicing efficiency. Further modifications include the possible methylation of the 2'-hydroxyl groups of the first two ribose sugars of the 5'

end of the mRNA. The capping process is essential to creating mature mRNAs that can then undergo translation.

The highly regulated capping process begins by RNA terminal phosphatase removing the gamma phosphate on the 5' end of mRNA. Then, guanylyl transferase cleaves a pyrophosphate off GTP and adds GMP to the terminal phosphates resulting in a 5' to 5' triphosphate linkage. Finally, a methyl transferase adds a methyl group to the 7'-nitrogen of guanine. Capping is tightly controlled and necessary for creating mRNAs that will be translated into proteins.

1.2.4 Functional Roles for RNA Chemical Modifications

The level of RNA modification is sensitive to various metabolic stress conditions and developmental stages.[58] Growth conditions and the environment can affect tRNA modifications both quantitatively and qualitatively, e.g. bacteria growing under starvation conditions for certain amino acids or iron leads to under-modification of tRNA.[63] The link between the synthesis of modified nucleosides in tRNA and metabolism has been suggested to be a regulatory device and tRNA modification as a "biological sensor."

However, the functional roles of modifications have been difficult to determine because unmodified RNA has subtle (if any) phenotype changes. For most tRNAs, unmodified versions have almost the same efficiency as the modified RNAs in accepting amino acids. However, a few unmodified tRNAs are severely affected in the aminoacylation reaction since these tRNAs lack modified nucleosides in the wobble position or 3' to the anticodon that is necessary for the cognate aminoacyl-ligase recognition.[64] While most completely unmodified tRNAs are able to accept amino acids *in vitro*, under certain conditions such tRNAs show considerable changed kinetics of aminoacylation because unmodified tRNA does not adopt the native conformation.

Therefore, modifications change the tRNA structure in a way that may influence the efficiency of the aminoacylation reaction and counteract mischarging. But there are examples, such as the unmodified *E. coli* tRNAs specific for methionine and valine, that are not mischarged suggesting that modified nucleosides are not always fidelity markers for tRNA identity. Generally, modified nucleosides improve the performance of the tRNA in the various processes of cellular metabolism.

While taken individually, modified nucleosides are rarely indispensable and can be absent in functional tRNA, but their complete absence may be deleterious. In *E. coli*, no modified nucleosides were shown to be essential for viability, but the lack of certain modifying enzymes can lead to lethality.[65] In *S. cerevisiae*, three tRNA modifying enzymes (Gcd10p/Gcd14p, Tad2p/Tad3p and Thg1p) that modify m^1A58,[66] I34[67] and $tRNA^{His}$ G_{-1} (guanine nucleotide to the 5'-end of $tRNA^{His}$)[68] are known to be essential. Deficiency of modified nucleosides can, therefore, lead to reduced translation efficiency and increased translation errors, which will affect gene expression regulation and cell metabolism.[64]

The possibility exists that certain modified nucleosides are important in other cellular processes in which tRNA molecules are involved (such as regulation of gene expression, links to other metabolic processes like synthesis of porphyrines or cell wall, cell differentiation, development of certain diseases like cancer or certain forms of auto-immunity).[57] The structural and thermodynamic effects of nucleoside modifications depend on the structural context, and can extend beyond the site of modification.

Although the regulatory effect of hypomodification may occur at several places in an mRNA (presence of specific structures, sequences prone to frameshifting, etc.), the 5' end seems to be especially sensitive to small translational aberrations. Ribosomes stalled early on an

mRNA due to modification deficiency will indirectly influence translation initiation of that mRNA, resulting in a lower degree of translation. There is a preferential usage of rare codons within the 25 first codons of mRNAs.[58] This codon bias early in the mRNA is critical for the efficiency of translation. Therefore, some mRNAs may have sites in the beginning that are sensitive to the modification status of the tRNA reading these sites.

1.3 The Discovery of Modified Nucleosides

Before 1948, nucleic acids were thought to contain only the four canonical bases, without any modification. Within the next decade, 5'-methylcytosine was found in DNA hydrosylates[69] and what later became called pseudouridine was characterized as the fifth RNA base.[70] The development of methods for purifying and sequencing different types of RNAs from diverse organisms in the three domains of life have led to the identification of more than a hundred modified nucleosides.[71]

Traditionally, novel modified nucleosides are discovered by digesting isolated nucleic acids with nuclease followed by purifying and characterizing the fractions by chromatography.[71] Indeed, the first modified nucleosides were found using this method.[69,70] In general, researchers subjected RNA to nuclease digestion and phosphatase followed by labeling to incorporate a radiolabel into the mononucleotides or short oligonucleotides. They subjected the digested and labeled nucleotides to thin-layer chromatography (TLC), purified the spots and characterized the products. Comparison with authentic standards confirmed their identities.

As analytical methods advanced, high-performance liquid chromatography (HPLC) and mass spectrometry replaced TLC for the identification of novel nucleoside modifications.[72-74] The resolution achieved with the liquid chromatography system coupled with the high sensitivity

of the mass spectrometer is optimal for the discovery and characterization of modified nucleosides. Furthermore, experiments such as tandem MS and collision induced dissociation in the LC/MS system allows for more facile structural elucidation. The procedure for analyzing digested nucleic acids is also streamlined due to the removal of the radiolabeling step. Thus currently, electrospray ionization (ESI) LC/MS is the method of choice to characterize modified nucleosides.

While the technology has improved for detecting and characterizing novel modifications on nucleic acids, many of the studies are targeted toward specific classes of RNA. For example, researchers have isolated tRNAs, rRNA, mRNAs as well as sRNAs to digest and investigate potential nucleoside modifications. Therefore, an unbiased screening approach that probes the modified nucleosides from total RNA would present a more complete picture of the chemical diversity of cellular RNA. Developments in LC/MS would aid in this method to sensitively and accurately detect novel nucleoside modifications that are present in cellular RNA.

1.4 The Functional Capacity of Small Molecule-Nucleic Acid Conjugates

The RNA world hypothesis proposes that at the origin of life, the first self-replicating system contained only RNA.[75] During this time, RNA would serve as both the genetic material and the principal cellular enzyme. Due to the discovery of catalytic RNAs, Gilbert postulated that RNA may have been capable of catalyzing the synthesis of themselves.[75] Based on the mechanism of the self-splicing intron, RNAs may have been able to splice itself out of and back into an appropriate nucleotide sequence. Thus, in an RNA world, RNAs may use this method for recombination and introduction of mutations against a background of replicating RNA molecules. RNA is ideally suited for this because it can carry replicable genetic information

while having the flexibility of assuming a wide variety of secondary and tertiary structures. The RNA can acquire additional enzymatic functions by using metal ions, amino acids and small molecule cofactors such as nicotinamide adenine dinucleotide and flavin mononucleotide.[75]

To form the lipid bilayer of the cell membrane in an RNA world, there needed to be a mechanism to make carbon-carbon bonds for the oligomerization process leading to polyketide and polyprenoid structures.[76] Tauer and colleagues have also hypothesized that RNA participated in aldol, Claisen and transmethylation reactions in addition to porphyrin biosynthesis.[77] Scott proposed an RNA-templated mechanism for the synthesis of polyketide, and by extension, fatty acid and polyprenoid biosynthesis. Hybridization of misacylated tRNA with acetyl and malonyl equivalents to a short RNA would allow the 3'-esters to engage in the necessary Claisen chemistry of polyketide synthesis. If the monomeric nucleosides were charged with 3'-phosphorylated RNA, the orientation would allow the formation of terpenoids, thereby providing a second avenue for production of lipids in the primitive organism.[76]

Indeed, small molecules linked to nucleic acids have demonstrated a wide range of chemistries *in vitro*. Specifically, functionalized nucleic acids can participate in a variety of chemical reactions between duplex DNA-linked reagents in DNA-templated organic synthesis (DTS).[78-80] The Watson-Crick base pairing of DNA-linked reagents significantly increases the effective molarity of reactivity while maintaining partner specificity to allow numerous DTS reactions in a single solution. DTS has been used to synthesize small molecules,[78,80,81] discover new reactions[82,83] and generate macrocycle libraries.[84] Given the general *in vitro* utility of DNA-linked small molecules, cellular systems may employ functionalized nucleic acids.

Primitive organisms may have used a RNA-based mechanism for carbon-carbon bond formation that led to lipids and porphyrins.[76] Nucleic acid-linked small molecules may be a

remnant from the hypothesized RNA world, where RNA catalyzed their own replication.[75] As in DTS, linked reagents may have broadened the scope of RNA mediated catalysis and improve the fidelity of information transfer, leading to evolutionary preservation. An investigation of small molecule-RNA conjugates in extant organisms may yield insights into the RNA world and the earliest biochemistry.

1.5 Thesis Overview

The known chemical diversity of natural RNA has remained limited despite a growing number of elucidated biological roles for RNA. Canonical polyribonucleotides, 3'-aminoacylated tRNAs,[7] modified nucleobases,[85] and (in eukaryotes) 5'-capped mRNAs[86,87] are the major known chemical components of natural RNA. In contrast, the more recent discovery of ribozymes,[50] riboswitches,[88] microRNAs (miRNAs),[89] small interfering RNAs (siRNAs),[90] Piwi-interacting RNAs (piRNAs),[91] small nuclear RNAs (snRNAs),[92] CRISPR sRNAs,[93] RNA transcriptional regulators,[94] and long non-coding RNAs[95,96] have greatly expanded the known functional roles of natural RNA beyond those described in the central dogma. As a result of these discoveries, RNA is now known to play a variety of catalytic, regulatory, and defensive roles in living systems.

Researchers have previously speculated that early biotic systems carried out biochemical reactions using small molecule-RNA conjugates, perhaps in the form of RNA-templated chemistries.[77,97-102] Recent studies involving artificial DNA-templated chemistries[79,84,103,104] also highlight the unusual functional capabilities of small molecule-nucleic acid conjugates. These observations collectively led us to speculate that small molecule-RNA conjugates beyond those previously described may exist in modern cells.[105]

By exploiting the improved resolution of ultra-performance liquid chromatography (UPLC) and the increased sensitivity of current MS methods, we can detect and characterize small molecule-RNA conjugates that are present as low as 30 copies per bacterial cell, representing a ~200-fold increase in sensitivity over previously described methods. Furthermore, the known examples of modified nucleosides have predominantly been discovered through serendipity or through a focused, hypothesis-driven approach. In principle, a broad, unbiased analysis of all small molecule-RNA conjugates in a cell may reveal a much larger set of such species. We are studying the entire set of nucleic acids that are present in a cell, rather than isolating a specific class or type of RNAs such as tRNA or rRNA. We also do not require a specific type of chemical linkage or reactivity for detection.

We have developed and implemented a method that in principle enables the detection of any small molecule-RNA conjugate regardless of its chemical structure. Our method uses enzymatic digestion of total cellular RNA to mononucleotides and comparative high-resolution liquid chromatography and mass spectrometry (LC/MS) coupled with computational analysis to identify non-canonical masses that are putatively linked to cellular RNA. MS/MS fragmentation, isotope labeling and comparison with authentic standards are then used to elucidate the structures of small molecules derived from conjugation to biological RNAs. Using this new method, we identified nicotinamide adenine dinucleotide (NAD) as covalently attached to unknown cellular RNA(s) in bacteria and some eukaryotes. We also characterized basic features of NAD-RNA conjugates from these organisms, including the nature of the NAD-RNA linkage, the inability of *E. coli* RNA polymerase to initiate transcription with NADH *in vitro*, and the size distribution of NAD-linked RNAs. Our approach has the potential to provide a more complete understanding of the chemical diversity of cellular RNA and add to the biological functions of RNA.

1.6 References

1 Palade, G. E. & Siekevitz, P. PANCREATIC MICROSOMES. *The Journal of Biophysical and Biochemical Cytology* **2**, 671-690, doi:10.1083/jcb.2.6.671 (1956).

2 Palade, G. E. & Porter, K. R. STUDIES ON THE ENDOPLASMIC RETICULUM. *The Journal of Experimental Medicine* **100**, 641-656, doi:10.1084/jem.100.6.641 (1954).

3 Porter, K. R. & Palade, G. E. STUDIES ON THE ENDOPLASMIC RETICULUM. *The Journal of Biophysical and Biochemical Cytology* **3**, 269-300, doi:10.1083/jcb.3.2.269 (1957).

4 Siekevitz, P. & Palade, G. E. A Cytochemical Study on the Pancreas of the Guinea Pig. *The Journal of Biophysical and Biochemical Cytology* **4**, 309-318, doi:10.1083/jcb.4.3.309 (1958).

5 Palade, G. E. & Siekevitz, P. LIVER MICROSOMES. *The Journal of Biophysical and Biochemical Cytology* **2**, 171-200, doi:10.1083/jcb.2.2.171 (1956).

6 Zamecnik, P. C. & Keller, E. B. RELATION BETWEEN PHOSPHATE ENERGY DONORS AND INCORPORATION OF LABELED AMINO ACIDS INTO PROTEINS. *Journal of Biological Chemistry* **209**, 337-354 (1954).

7 Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **231**, 241-257 (1958).

8 Crick, F. H. On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138-163 (1958).

9 Brenner, S., Jacob, F. & Meselson, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **190**, 576-581 (1961).

10 Bishop, J. M. *et al.* The low molecular weight RNAs of Rous sarcoma virus: II. The 7 S RNA. *Virology* **42**, 927-937 (1970).

11 Walker, T. A., Pace, N. R., Erikson, R. L., Erikson, E. & Behr, F. The 7S RNA Common to Oncornaviruses and Normal Cells is Associated with Polyribosomes. *Proceedings of the National Academy of Sciences* **71**, 3390-3394 (1974).

12 Zieve, G. & Penman, S. Small RNA species of the HeLa cell: Metabolism and subcellular localization. *Cell* **8**, 19-31 (1976).

13 Walter, P., Ibrahimi, I. & Blobel, G. n. Translocation of Proteins across the Endoplasmic Reticulum I. Signal Recognition Protein (SRP) Binds to In-vitro-Assembled Polysomes Synthesizing Secretory Protein. *The Journal of Cell Biology* **91**, 545-550 (1981).

14 Walter, P. & Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**, 691-698 (1982).

15 Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. A possible precursor of immunoglobulin light chains. *Nature: New biology* **239**, 117-120 (1972).

16 Keiler, K. C. Biology of trans-Translation. *Annual Review of Microbiology* **62**, 133-151, doi:10.1146/annurev.micro.62.081307.162948 (2008).

17 Ray, B. K. & Apirion, D. Characterization of 10S RNA: A new stable RNA molecule from <i>Escherichia coli</i&gt. *Molecular and General Genetics MGG* **174**, 25-32, doi:10.1007/bf00433301 (1979).

18 Tyagi, J. S. & Kinger, A. K. Identification of the 10Sa RNA structural gene of Mycobacterium tuberculosis. *Nucleic Acids Research* **20**, 138, doi:10.1093/nar/20.1.138 (1992).

19 Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proceedings of the National Academy of Sciences* **91**, 9223-9227 (1994).
20 Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. H. A. mut-7 of C. elegans, Required for Transposon Silencing and RNA Interference, Is a Homolog of Werner Syndrome Helicase and RNaseD. *Cell* **99**, 133-141 (1999).
21 Tabara, H. *et al.* The rde-1 Gene, RNA Interference, and Transposon Silencing in C. elegans. *Cell* **99**, 123-132 (1999).
22 Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11**, 1017-1027 (2001).
23 Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366, doi:http://www.nature.com/nature/journal/v409/n6818/suppinfo/409363a0_S1.html (2001).
24 Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. *Cell* **101**, 25-33 (2000).
25 Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development* **15**, 188-200, doi:10.1101/gad.862301 (2001).
26 Bass, B. L. Double-Stranded RNA as a Template for Gene Silencing. *Cell* **101**, 235-238 (2000).
27 Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215-233 (2009).
28 Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906 (2000).
29 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
30 Lee, R. C. & Ambros, V. An Extensive Class of Small RNAs in Caenorhabditis elegans. *Science* **294**, 862-864, doi:10.1126/science.1065329 (2001).
31 Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, 858-862, doi:10.1126/science.1065062294/5543/858 [pii] (2001).
32 Brodersen, P. *et al.* Widespread Translational Inhibition by Plant miRNAs and siRNAs. *Science* **320**, 1185-1190, doi:10.1126/science.1159151 (2008).
33 Seto, A. G., Kingston, R. E. & Lau, N. C. The Coming of Age for Piwi Proteins. *Molecular Cell* **26**, 603-609 (2007).
34 Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246-258 (2011).
35 Brantl, S. Antisense-RNA regulation and RNA interference. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1575**, 15-25 (2002).
36 Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **81**, null, doi:doi:10.1146/annurev-biochem-051410-092902 (2012).

37 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:321/5891/960 [pii]10.1126/science.1159689 (2008).
38 Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem Biol* **9**, 1043, doi:S1074552102002247 [pii] (2002).
39 Mironov, A. S. *et al.* Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**, 747-756 (2002).
40 Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952-956 (2002).
41 Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences* **99**, 15908-15913, doi:10.1073/pnas.212628899 (2002).
42 Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences* **29**, 11-17 (2004).
43 Nou, X. & Kadner, R. J. Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proceedings of the National Academy of Sciences* **97**, 7190-7195, doi:10.1073/pnas.130013897 (2000).
44 Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends in Genetics* **15**, 439-442 (1999).
45 Stormo, G. D. & Ji, Y. Do mRNAs act as direct sensors of small molecules to control their expression? *Proceedings of the National Academy of Sciences* **98**, 9465-9467, doi:10.1073/pnas.181334498 (2001).
46 Gold, L. *et al.* From oligonucleotide shapes to genomic SELEX: Novel biological,Äâregulatory,Äâloops. *Proceedings of the National Academy of Sciences* **94**, 59-64 (1997).
47 Woese, C. *The genetic code*. (Harper and Row, New York, 1967).
48 Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157 (1982).
49 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
50 Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228, doi:10.1038/418222a418222a [pii] (2002).
51 Perrotta, A. T., Shih, I.-h. & Been, M. D. Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme. *Science* **286**, 123-126, doi:10.1126/science.286.5437.123 (1999).
52 Santoro, S. W., Joyce, G. F., Sakthivel, K., Gramatikova, S. & Barbas, C. F. RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. *Journal of the American Chemical Society* **122**, 2433-2439, doi:10.1021/ja993688s (2000).
53 Wyatt, G. R. Occurrence of 5-methylcytosine in nucleic acids. *Nature* **166**, 237-238 (1950).
54 Limbach, P. A., Crain, P. F. & McCloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22**, 2183-2196 (1994).
55 Gott, J. M. in *Methods in enzymology* (Academic Press/Elsevier, 2007).
56 Czerwoniec, A. *et al.* MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Research* **37**, D118-D121, doi:10.1093/nar/gkn710 (2009).
57 Motorin, Y. & Grosjean, H. in *eLS* (John Wiley & Sons, Ltd, 2001).

58 Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-351, doi:S0968000402021096 [pii] (2002).
59 Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **289**, 920-930, doi:10.1126/science.289.5481.920 (2000).
60 Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 √Ö Resolution. *Science* **289**, 905-920, doi:10.1126/science.289.5481.905 (2000).
61 Harms, J. *et al.* High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. *Cell* **107**, 679-688 (2001).
62 Yusupov, M. M. *et al.* Crystal Structure of the Ribosome at 5.5 √Ö Resolution. *Science* **292**, 883-896, doi:10.1126/science.1060089 (2001).
63 Kitchingman, G. R. & Fournier, M. J. Modification-deficient transfer ribonucleic acids from relaxed control Escherichia coli: structures of the major undermodified phenylalanine and leucine transfer RNAs produced during leucine starvation. *Biochemistry* **16**, 2213-2220, doi:10.1021/bi00629a027 (1977).
64 Grosjean, H. in *Topics in current genetics,; 12;* (New York, 2005).
65 Persson, B. C., Gustafsson, C., Berg, D. E. & Bj√∂rk, G. R. The gene for a tRNA modifying enzyme, m5U54-methyltransferase, is essential for viability in Escherichia coli. *Proceedings of the National Academy of Sciences* **89**, 3995-3998 (1992).
66 Anderson, J. *et al.* The essential Gcd10p,ÄìGcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes & Development* **12**, 3650-3662, doi:10.1101/gad.12.23.3650 (1998).
67 Gerber, A. P. & Keller, W. An Adenosine Deaminase that Generates Inosine at the Wobble Position of tRNAs. *Science* **286**, 1146-1149, doi:10.1126/science.286.5442.1146 (1999).
68 Gu, W., Jackman, J. E., Lohan, A. J., Gray, M. W. & Phizicky, E. M. tRNAHis maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5,Ä≤ end of tRNAHis. *Genes & Development* **17**, 2889-2901, doi:10.1101/gad.1148603 (2003).
69 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *The Biochemical journal* **55**, 774-782 (1953).
70 Cohn, W. E. & Volkin, E. Nucleoside-5[prime]-Phosphates from Ribonucleic Acid. *Nature* **167**, 483-484 (1951).
71 Grosjean, H. Vol. 12 *Topics in Current Genetics* (ed Henri Grosjean) 1-22 (Springer Berlin / Heidelberg, 2005).
72 Buck, M., Connick, M. & Ames, B. N. Complete analysis of tRNA-modified nucleosides by high-performance liquid chromatography: The 29 modified nucleosides of Salmonella typhimurium and Escherichia coli tRNA. *Analytical Biochemistry* **129**, 1-13 (1983).
73 Klawitter, J., Schmitz, V., Klawitter, J., Leibfritz, D. & Christians, U. Development and validation of an assay for the quantification of 11 nucleotides using LC/LC,Äìelectrospray ionization,ÄìMS. *Analytical Biochemistry* **365**, 230-239 (2007).
74 Tuytten, R., Lemière, F., Dongen, W. V., Esmans, E. L. & Slegers, H. Short capillary ion-pair high-performance liquid chromatography coupled to electrospray (tandem) mass spectrometry for the simultaneous analysis of nucleoside mono-, di- and triphosphates.

Rapid Communications in Mass Spectrometry **16**, 1205-1215, doi:10.1002/rcm.704 (2002).
75 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618-618 (1986).
76 Scott, A. I. How were porphyrins and lipids synthesized in the RNA world? *Tetrahedron Letters* **38**, 4961-4964 (1997).
77 Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci U S A* **86**, 7054-7058 (1989).
78 Gartner, Z. J., Kanan, M. W. & Liu, D. R. Multistep small-molecule synthesis programmed by DNA templates. *Journal of the American Chemical Society* **124**, 10304-10306 (2002).
79 Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. & Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **431**, 545-549 (2004).
80 Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *Journal of the American Chemical Society* **123**, 6961-6963 (2001).
81 Zev J. Gartner, M. W. K. D. R. L. Expanding the Reaction Scope of DNA-Templated Synthesis. *Angewandte Chemie International Edition* **41**, 1796-1800 (2002).
82 Rozenman, M. M., Kanan, M. W. & Liu, D. R. Development and Initial Application of a Hybridization-Independent, DNA-Encoded Reaction Discovery System Compatible with Organic Solvents. *Journal of the American Chemical Society* **129**, 14933-14938 (2007).
83 Momiyama, N., Kanan, M. W. & Liu, D. R. Synthesis of acyclic alpha,beta-unsaturated ketones via Pd(II)-catalyzed intermolecular reaction of alkynamides and alkenes. *Journal of the American Chemical Society* **129**, 2230-2231 (2007).
84 Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *J Am Chem Soc* **130**, 15611-15626, doi:10.1021/ja805649f (2008).
85 Dunin-Horkawicz, S. *et al.* MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* **34**, D145-149, doi:34/suppl_1/D145 [pii]10.1093/nar/gkj084 (2006).
86 Wei, C. M. & Moss, B. Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci U S A* **72**, 318-322 (1975).
87 Furuichi, Y. & Miura, K. A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253**, 374-375 (1975).
88 Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463, doi:10.1038/nrm1403nrm1403 [pii] (2004).
89 Chen, K. & Rajewsky, N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**, 93-103, doi:nrg1990 [pii]10.1038/nrg1990 (2007).
90 Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nat Rev Genet* **6**, 24-35, doi:nrg1500 [pii]10.1038/nrg1500 (2005).
91 Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21**, 2300-2311, doi:21/18/2300 [pii]10.1101/gad.1564307 (2007).
92 Patel, S. B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**, 6482-6493, doi:gkn658 [pii]10.1093/nar/gkn658 (2008).

93 Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186, doi:nrmicro1793 [pii]10.1038/nrmicro1793 (2008).
94 Storz, G., Altuvia, S. & Wassarman, K. M. An abundance of RNA regulators. *Annu Rev Biochem* **74**, 199-217, doi:10.1146/annurev.biochem.74.082803.133136 (2005).
95 Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic Acids Res* **37**, D122-126, doi:gkn617 [pii]10.1093/nar/gkn617 (2009).
96 Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29, doi:15/suppl_1/R17 [pii]10.1093/hmg/ddl046 (2006).
97 Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci U S A* **96**, 5470-5475 (1999).
98 Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387-390, doi:10.1038/35053176 (2001).
99 Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J Mol Evol* **46**, 18-36 (1998).
100 Visser, C. M. & Kellogg, R. M. Bioorganic chemistry and the origin of life. *J Mol Evol* **11**, 163-168 (1978).
101 White, H. B., 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101-104 (1976).
102 Calderone, C. T. & Liu, D. R. Nucleic-acid-templated synthesis as a model system for ancient translation. *Curr Opin Chem Biol* **8**, 645-653, doi:S1367-5931(04)00127-9 [pii]10.1016/j.cbpa.2004.09.003 (2004).
103 Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601-1605, doi:10.1126/science.11026291102629 [pii] (2004).
104 Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew Chem Int Ed Engl* **43**, 4848-4870, doi:10.1002/anie.200400656 (2004).
105 Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A* **106**, 7768-7773, doi:0900528106 [pii]10.1073/pnas.0900528106 (2009).

Chapter Two:

The Development and Application of a Nuclease-Based Screen that Reveals NAD-Linked RNA

Ye Grace Chen, Walter Kowtoniuk, Yinghua Shen, David R Liu

Ye Grace Chen and Walter Kowtoniuk developed the nuclease-based screen. Ye Grace Chen and Yinghua Shen developed the soft ionization LC/MS method. Walter Kowtoniuk carried out the *in vitro* transcription initiation by *E. coli* polymerase experiment. Ye Grace Chen conducted and analyzed all of the other experiments described.

Adapted from "LC/MS analysis of cellular RNA reveals NAD-linked RNA."
Nature Chemical Biology (2009) 5(12): 879-881.

2.1 Introduction

The known chemical diversity of cellular RNA remains modest compared with the rapidly growing diversity of RNA's known biological roles. In this chapter, we describe the development and application of a general method to detect small molecule-linked RNAs that does not rely on the chemical reactivity of the small molecule, and thus in principle enables the detection of any cellular small molecule-RNA conjugate. We validated the method by detecting known tRNA and rRNA modifications in *Escherichia coli* and *Streptomyces venezuelae* RNA. When applied to *E. coli* and *S. venezuelae* RNA, the method revealed NAD-linked RNA, in addition to a number of other previously unreported putative small molecule-RNA conjugates. Subsequent experiments to characterize NAD-linked RNA revealed that the NAD group is attached to unknown *E. coli* and *S. venezuelae* RNA(s) at the 5' terminus and cannot be installed *in vitro* through aberrant transcriptional initiation by *E. coli* RNA polymerase. Strikingly, quantitation using authentic standards indicates that NAD-linked RNA is present at ~3,000 copies per cell, a level comparable to that of several of the more abundant aminoacylated tRNAs. These results provide a new example of a biological small molecule-RNA conjugate, further support our conclusion that the chemical diversity of cellular RNA includes unanticipated classes of molecules, and highlight the value of the general chemical screening method developed in this work.

2.2 A More General Method for Detecting Cellular Small Molecule-RNA Conjugates

We previously developed two methods to detect cellular small molecule-RNA conjugates that depend on the chemical lability of the small molecule or nucleotide to which the small molecule is attached.[105] We compared total RNA that was digested with nuclease, and then

treated with base or nucleophile to samples that were treated in conditions that would leave the nucleotide intact, and analyzed by reverse phase liquid chromatography/electrospray ionization-mass spectrometry (LC/ESI-MS). By treating half of the sample with either base or nucleophile and the other half with conditions that should not affect RNA cleavage, we discovered base- or nucleophile-labile small molecules that are linked to RNA.

In order to generalize the screen that detects novel small molecule modifications on RNA, we determined that the samples for comparative analysis could be derived from (*i*) RNA treated with a nuclease enzyme that cleaved the RNA to mononucleotides, and (*ii*) RNA treated with heat-inactivated nuclease that left the RNA intact but contained all other small molecules introduced or generated during RNA isolation and sample processing (Fig. 2.1). Non-canonical species more abundant in the active nuclease-treated sample than in the inactive nuclease-treated sample could then be identified as possible novel small molecule-RNA conjugates, regardless of their chemical lability. Based on the sensitivity of our current LC/MS methods, this method would be able to detect ~0.5 pmol of a given nucleotide, corresponding to detection of modifications that are present at ~30 copies per bacterial cell. This sensitivity represents a ~200-fold improvement on previous efforts to study modified nucleotides, which usually focused on specific fractions of the cellular RNA (e.g., tRNAs) and were limited to studying nucleotides at high abundance levels of greater than or equal to ~2%; in *E. coli* this abundance level corresponds to ~4,400 copies per cell.[106,107]

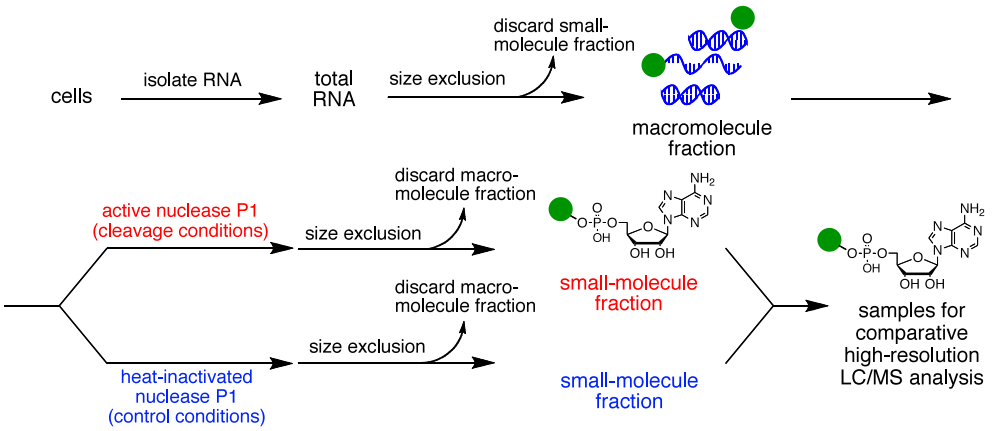


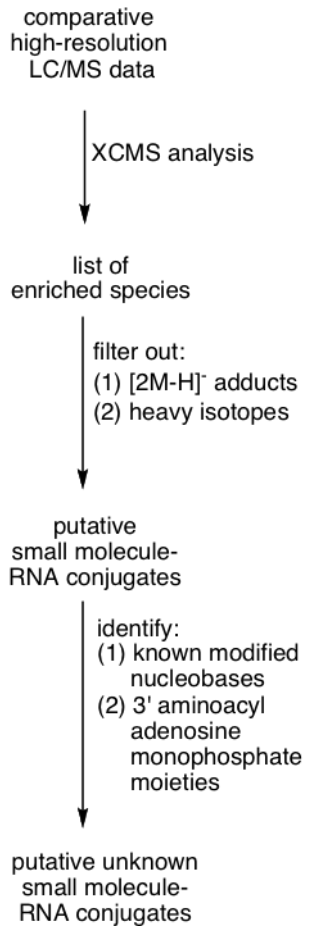
Figure 2.1. The general method for biological small molecule-RNA conjugate discovery developed in this work. The paired samples for comparative LC/MS analysis were generated by treatment with active nuclease P1 versus treatment with heat-inactivated nuclease P1 under otherwise identical conditions.

2.3 Validation Using Aminoacylated tRNAs and Known Nucleoside Modifications

The methods outlined above were implemented and applied to *Halobacterium salinarum, Saccharomyces cerevisiae*, *E. coli* and *S. venezuelae* RNA (Fig. 2.1). Whole cellular RNA was subjected to size-exclusion chromatography and the macromolecular fraction (greater than ~2,500 Da) was divided into two halves. One half was treated for 40 min at 37 °C with nuclease P1, an endonuclease that cleaves RNA to generate mononucleotides with a 3' hydroxyl and a 5' phosphate.[108] The second half was treated with heat-inactivated nuclease P1 under otherwise identical conditions. Both samples were subjected to size-exclusion chromatography again and the small-molecule fraction from each was retained. The two samples were then analyzed by LC/MS. Peaks with corresponding retention times containing species with similar mass:charge ratios (*m/z*) from the two samples were computationally paired, and their relative abundances were calculated.[105] Species with greater abundance in the active nuclease-treated sample than in the inactive heat-inactivated nuclease sample were considered candidate nucleotides for further study (Fig. 2.1).

Compared with our previous methods, this more general approach places a greater burden on the downstream processing of the resulting comparative LC/MS data. We therefore computationally filtered out (Fig. 2.2) canonical mono- and di-ribonucleotides, all known modified nucleotides including aminoacylated AMPs, all known base-modified tRNA and rRNA nucleotides, and the common ionization fragments and isotope peaks from all of these species. This set of known species is significantly larger than the set previously used in the chemical cleavage methods due to the more comprehensive nature of this approach.

a



b

3'-aminoacyl adenosine monophosphate	expected $[M-H]^-$ *m/z*
alanine	417.0929
arginine	502.1569
asparagine	460.0987
aspartic acid	461.0827
cysteine	449.0650
glutamic acid	475.0984
glutamine	474.1144
glycine	403.0772
histidine	483.1147
isoleuicine	459.1399
leucine	459.1399
lysine	474.1507
methionine	477.0963
phenylalanine	493.1242
proline	443.1085
serine	433.0878
threonine	447.1035
tryptophan	532.1351
tyrosine	509.1191
valine	445.1242

known nucleotide monophosphate	expected $[M-H]^-$ *m/z*	known nucleotide monophosphate	expected $[M-H]^-$ *m/z*
galQ	650.1716	$preQ_0$	386.0507
manQ		mnm^5s^2U	382.0480
o_2yW	619.1407	cm^5U	381.0341
OHyW	603.1458	ncm^5U	380.0501
yW	587.1508	ncm^5Um	
ms^2hn^6A	551.0967	ac^4Cm	378.0708
ms^2t^6A	537.0810	Gm	376.0664
OHyW*	531.1246	m^1G	
hn^6A	505.1090	m^2G	
m^6t^6A		m^7G	
oQ	504.1137	m^1Im	375.0711
Gr(p)	494.0930	m^1Am	374.0871
t^6A	491.0933	$m^6{}_2A$	
Q	488.1188	m^6Am	
Ar(p)	478.0981	nm^5s^2U	368.0323
ms^2io^6A	476.1010	mnm^5U	366.0708
tm^5s^2U	476.0204	$m^4{}_2Cm$	364.0915
ms^2i^6A	460.1061	ac^4C	364.0551
tm^5U	460.0433	f^5Cm	
k^2C	450.1395	GMP	362.0507
g^6A	447.0671	Im	361.0555
m^1acp^3Y	438.0919	m^1I	
inm^5s^2U	436.0949	Am	360.0715
inm^5Um	434.1334	m^1A	
io^6A	430.1133	m^2A	
mnm^5se^2U	429.9924	m^6A	
mimG	428.0977	mo^5U	353.0392
$cmnm^5s^2U$	426.0378	s^2Um	
acp^3U	424.0763	m^5s^2U	353.0214
$cmnm^5Um$		hm^5C	352.0551
inm^5U	420.1177	m^3Um	351.0599
i^6A	414.1184	m^5Um	
imG	414.0820	m^4Cm	350.0759
imG2		m^5Cm	
$mcmo^5U$	411.0446	f^5C	350.0395
$mcmo^5U$		I	347.0398
mcm^5s^2U	411.0269	AMP	346.0558
$cmnm^5U$	410.0606	m^5D	339.0599
mcm^5Um	409.0654	ho^5U	339.0235
ms^2m^6A	406.0592	s^2U	339.0058
$m^2{}_2Gm$	404.0977	s^4U	
$m^2{}_7Gm$		s^2C	338.0217
G^+	403.0773	m^3U	337.0442
imG-14	400.0664	m^3Y	
chm^5U	397.0290	m^5U	
cmo^5U		Um	
mcm^5U	395.0497	Cm	336.0602
m^1Gm	390.0820	m^3C	
$m^2{}_2G$		m^5C	
$m^2{}_7G$		D	325.0442
m^2Gm		UMP	323.0286
$preQ_1$		Y	
$m^6{}_2Am$	388.1028	CMP	322.0446
ac^6A	388.0664		

Figure 2.2. Computational filtering to identify putative unknown small molecule-RNA conjugates. (**a**) Following XCMS analysis,[109] MS adducts and isotopic species were filtered out to generate a list of putative small molecule-RNA conjugates. $[2M-H]^-$ adducts were filtered out by removing the heavier of any two peaks separated in *m/z* value by M Da with the same retention time. Heavy isotope species were filtered out by removing any peak with a corresponding peak 1.00627 Da (the mass of a neutron) lower in *m/z* with identical retention

Figure 2.2. **(continued):** times. From the resulting filtered list we identify known modified nucleobases were identified, shown in (b). Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

To validate the method, we processed whole *E. coli* and *S. venezuelae* cellular RNA and searched for species enriched in the active nuclease sample that had masses and retention times consistent with those of amino acid-linked adenosine monophosphates and nucleobase modifications known to exist in bacteria. Of the 20 major 3'-aminoacyl adenosine monophosphates, 16 (80%) were detected as enriched at least 2-fold in the active nuclease-treated samples compared with the heat-inactivated nuclease-treated samples in *E. coli* (Fig. 2.3a). In addition, 31 species enriched at least 2-fold were observed with masses consistent with the masses of known RNA nucleobase modifications (Fig. 2.3b). In *S. venezuelae*, 16 aminoacyl adenosine monophosphates (Fig 2.4a) and 26 known RNA nucleobase modifications were detected as enriched at least 2-fold. An enrichment threshold of 2-fold represented a reasonable balance between detecting the maximum number of known 3'-aminoacyl adenosine monophosphates and nucleobase modifications, and minimized the number of irreproducible false-positive signals resulting from experimental noise. We note that several of the detected nucleobase modifications such as queuosine, 5-hydroxyuridine, and 2-methylthio-N^6-isopentenyladenosine are not expected to be base- or nucleophile-labile, and were not detected by our previous methods that relied on base cleavage or nucleophile cleavage.[105] These results validate the ability of the nuclease versus heat-inactivated nuclease-based method to detect the presence of known small molecule-RNA conjugates from whole cellular RNA, including conjugates that are not chemically labile.

a

amino acid	expected [M-H]⁻ m/z	observed [M-H]⁻ m/z	nuclease/control enrichment
alanine	417.0929	417.0934	61
asparagine	460.0987	460.0977	6.7
glutamine	474.1144	474.1135	34
glycine	403.0772	403.0769	111
histidine	483.1147	483.1142	43
isoleuicine	459.1399	459.1392	746
leucine	459.1399	459.1387	537
lysine	474.1507	474.1499	53
methionine	477.0963	477.0959	123
phenylalanine	493.1242	493.1243	244
proline	443.1085	443.1086	17
serine	433.0878	433.0866	20
threonine	447.1035	447.1027	107
tryptophan	532.1351	532.1346	45
tyrosine	509.1191	509.1190	36
valine	445.1242	445.1229	728

b

RNA modification	expected [M-H]⁻ *m/z*	observed [M-H]⁻ *m/z*	retention time (min)	nuclease/control enrichment
m^6t^6A	505.1090	505.0983	8.1	150
hn^6A				
oQ	504.1137	504.1142	6.3	2.7
t^6A	491.0933	491.0971	2.0	13
Q	488.1333	488.1330	2.6	22
ms^2i^6A	460.1061	460.1089	13.7	142
$cmnm^5s^2U$	426.0378	426.0381	2.0	11
i^6A	414.1184	414.1173	10.3	67
$mcmp^5U$	411.0446	411.0434	3.9	206
cmo^5U	397.0290	397.0314	0.8	9.0
$m^2{}_2G$	390.0820	390.0802	6.2	109
$preQ_0$	386.0507	386.0493	0.8	1.7
mnm^5s^2U	382.0480	382.0498	1.6	1.5
m^1G	376.0664	376.0678	1.6	88
m^2G			1.8	3.9
m^7G			2.0	5.1
Gm			2.6	7.9
$m^6{}_2A$	374.0871	374.0878	7.0	14
ac^4C	364.0551	364.0543	1.2	64
m^1A	360.0715	360.0710	4.3	36
m^2A			5.3	54
m^6A			5.8	7
m^5s^2U			1.0	5.2
ho^5U	339.0235	339.0249	1.9	19
s^2U			1.6	5.6
s^2C	338.0217	338.0271	0.8	6.3
m^3U	337.0442	337.0442	1.3	19
Um			1.7	37
m^5U			1.9	13
m^3Y			2.4	5.6
m^3C	336.0602	336.0602	1.1	4.1
m^5C		336.0605	1.4	2.2
Cm		336.0611	1.7	3.5
D	325.0442	325.0461	0.8	12

Figure 2.3. (**a**) 3'-aminoacyl adenosine monophosphates from total *E. coli* RNA detected with the nuclease-based method. (**b**) Species detected by the nuclease digestion method that are consistent with known rRNA and tRNA nucleotide modifications from total *E. coli*. Although plausible identifications are shown based on high-resolution LC/MS data, we note that definitive identifications require comparison with authentic samples. Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

a

amino acid	expected $[M-H]^-$ m/z	observed $[M-H]^-$ m/z	nuclease/control enrichment
alanine	417.0929	417.0936	11
aspartic acid	461.0827	461.0816	7.8
glutamic acid	475.0984	475.0978	69
glutamine	474.1144	474.1149	20
glycine	403.07724	403.0770	41
isoleuicine	459.1399	459.1396	148
leucine	459.1399	474.1400	97
lysine	474.1507	474.1503	25
methionine	477.0963	477.0960	12
phenylalanine	493.1242	493.1238	62
proline	443.10854	443.1061	10
serine	433.0878	433.0887	10
threonine	447.1035	447.1035	33
tryptophan	532.1351	532.1348	28
tyrosine	509.1191	509.1196	16
valine	445.1242	445.1240	163

b

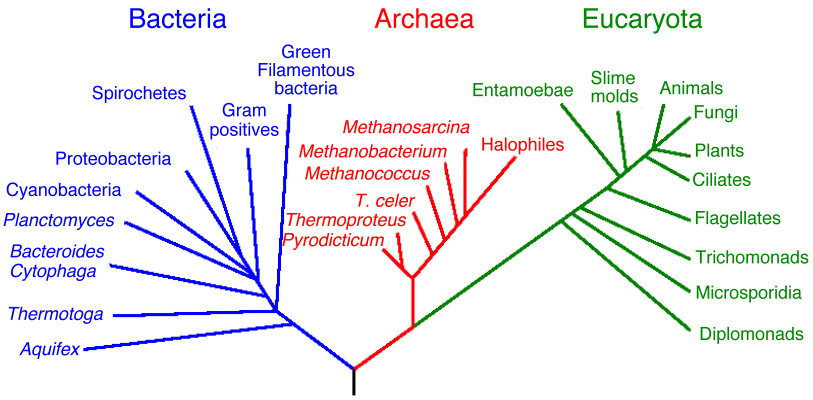
RNA modification	expected $[M-H]^-$ *m/z*	observed $[M-H]^-$ *m/z*	retention time (min)	nuclease/control enrichment
m^6t^6A	505.1090	505.0952	8.2	145
hn^6A				
Q	488.1333	491.0922	2.1	26
ms^2i^6A	460.1061	460.1085	13.7	29
$cmnm^5s^2U$	426.0378	426.0342	1.6	53
$cmnm^5Um$	424.0763	424.0680	1.0	3.0
inm^5U	420.1177	420.1159	6.0	11
i^6A	414.1184	414.1191	10.4	123
$m^2{}_2G$	390.0820	390.0852	6.2	99
m^7G			2.0	58
Gm			2.6	6.7
$m^6{}_2A$	374.0871	374.0892	7.0	83
nm^5s^2U	368.0323	368.0359	1.3	7.3
ac^4C	364.0551	364.0563	1.0	95
m^1A	360.0715	360.0739	5.0	3.5
m^2A			5.3	22
m^6A			5.7	26
m^5s^2U			1.0	5.2
m^3U	337.0442	337.0448	1.6	74
Um			2.0	6.7
m^5U			2.3	18
m^3Y			3.3	4.9
m^3C	336.0602	336.0572	1.4	49
m^5C			2.0	33
Cm			2.3	7.3
D	325.0442	325.0442	1.0	64

Figure 2.4. (**a**) 3'-aminoacyl adenosine monophosphates from total *S. venezuelae* RNA detected with the nuclease-based method. (**b**) Species detected with the nuclease digestion method that are consistent with known rRNA and tRNA nucleotide modifications from total *S. venezuelae*. Although plausible identifications are reported based on high-resolution LC/MS data, definitive identifications require additional comparisons with authentic samples of each modified nucleotide. Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

2.4 Application of the Nuclease-Based Screen to *Halobacterium salinarum* RNA

The study of small molecule-RNA conjugates may be particularly interesting in archaea due to their fundamental biological differences compared with bacteria and eukaryotes (Fig. 2.5). For example, archaea contain membrane lipids composed predominantly of isoprenoid glycerol diethers or diglycerol tetraethers, rather than glycerol esters, and their ribosomes have substantial rRNA structural differences compared to bacteria and eukaryotes.[110-112] Archaea also contain archaea-specific post-transcriptional RNA modifications,[54,113] including at least a dozen unique tRNA nucleobase modifications that are not known to exist in bacteria or eukaryotes.[85] Based on these differences, we anticipate that the application of our methods to archaea will continue to

expand our knowledge of the chemical diversity of biological RNA through examples that may not be present in bacteria or eukaryotes.



lipid chemistry	glycerol esters	glycerol ethers	glycerol esters
lipid composition	fatty acids	isoprenoids	fatty acids
cellular organelles	no	no	yes
chromosomes	1	multiple	multiple
RNA polymerases	1	3	3
5' capped mRNA	no	no	yes
polyadenylated mRNA	no	no	yes
known nucleoside modifications			
domain-unique	15	12	34
total	53	48	65

Figure 2.5. Phylogenetic tree and key biochemical difference among the three domains of life.[110,111,114]

Since we hypothesized that novel small molecule-RNA conjugates may exist in modern cells as evolutionary fossils, haloarchaea are of particular interest for study. Halophiles have been proposed as the earliest living cells to evolve, resulting from the concentration of both salts and organic compounds in an evaporitic environment giving rise to the primordial soup.[115,116] Of the archaean organisms, haloarchaea are among the most extensively studied due to their relative ease of laboratory culturing, transformability, genetic malleability, and their known genome sequences.[117] *Halobacterium salinarum* requires only high salt concentrations and aeration to grow, and the prevalent interest in understanding the functioning of cellular machinery under such salt concentrations has aided the development of methodologies that we can apply to

studying small molecule-RNA conjugates in *H. salinarum*.[117] For example, transformations and the creation of mutant strains in *H. salinarum* are now commonplace.[118,119] Transcriptome and proteome analyses have been also conducted for *H. salinarum*,[120-122] as well as a full genome sequencing,[114] and together these provide invaluable resources in our investigation of archaean small molecule-RNA conjugates.

We isolated total RNA from *H. salinarum* and treated half with nuclease P1 and the other half with heat-inactivated nuclease P1. Both sets of samples were subjected to size exclusion chromatography prior to LC/MS analysis and computation processing. In addition to detecting known modifications, we observed, on average, 11 of the 20 major aminoacyl-adenylates. The method also detected nine non-canonical species that were enriched at least 2-fold from the screens of *H. salinarum* total RNA (Fig. 2.6). We obtained initial MS/MS fragmentation data for each of the species.

m/z	rt (min)	Nuclease/Heat-Inactivated Nuclease Enrichment
536.070	2.947	115.530
394.104	2.954	66.775
898.381	3.354	22.949
395.105	2.964	14.867
572.109	4.540	13.996
759.628	23.151	6.595
559.045	3.024	5.996
446.132	7.286	5.748
421.164	17.005	5.505

Figure 2.6. The nuclease-based screen was applied to *H. salinarum* and nine novel small molecules were detected as reproducibly enriched.

While we were able to identify nucleotide components of the small molecule modifications of RNA, further structural characterization was elusive due to sample preparation limitations. In the initial studies, *H. salinarum* were cultured in baffled flasks for twelve to fourteen days after inoculation with a saturated starter culture. In order for the halophiles to

grow properly, a high oxygen supply needed to be balanced without overly vigorous lateral shaking since there have been reports of culture lysis under those conditions.[123] From the flask cultures, an average of ten mg of total RNA were extracted from eight liters of cells. Since each screen required at least three mg of total RNA and more for structure elucidation experiments, other methods of large scale culturing were investigated. *H. salinarum* were then grown in fermenters, which increased the cell density before growth was inhibited, and thus, the amount of total RNA extracted. However, the sample demands of the early phase screening coupled with the long doubling time *H. salinarum* led us to explore other organisms that might be more technically tractable for investigation of small molecule-RNA conjugates.

2.5 Application of the Nuclease-Based Screen to *Saccharomyces cerevisiae* RNA

Eukaryotes, such as yeast, are attractive species for our investigations, due to their ease of laboratory culturing and well-developed molecular biology protocols for their study.[124-127] Furthermore, preliminary evidence suggests the potential biological relevance of small molecule-RNA conjugates in eukaryotes. That the lack of certain nucleoside modifications leads to phenotypic changes in yeast, such as ceasing sporulation, indicates that the post-transcriptional modifications are involved in gene regulation or cellular signaling.[128] A comprehensive investigation of small molecule-RNA conjugates in *S. cerevisiae* using the nuclease-based screening method may therefore lead to the discovery of new biological pathways.[129] Thus, we applied the previously developed methods to the thorough search for cellular small molecule-RNA conjugates in the eukaryote *S. cerevisiae*. When *S. cerevisiae* total RNA was subjected to the nuclease-based screen, 23 unknown species was enriched at least 2-fold (Fig. 2.7). Two

independent biological replicates generated enrichment factors with a trial-to-trial correlation coefficient of 0.94.

m/z	rt (min)	Nuclease/Heat-Inactivated Nuclease Enrichment
459.250	8.306	122.818
917.430	9.152	79.308
415.220	10.570	76.325
781.355	5.632	70.300
721.321	5.301	58.258
490.228	5.459	44.339
902.408	7.336	36.603
723.294	5.641	36.015
830.439	10.553	29.353
749.361	6.946	25.578
446.111	5.658	25.279
696.281	5.846	16.857
682.254	5.275	16.815
585.513	6.471	12.554
493.243	9.334	12.053
509.239	7.587	11.315
520.409	8.918	10.949
810.541	10.981	6.594
636.442	16.306	5.852
631.417	14.826	4.374
980.699	11.999	3.455
608.408	14.861	3.365
690.964	12.043	3.198

Figure 2.7. The nuclease-based screen was applied to *S. cerevisiae* and twenty-three novel small molecules were detected as reproducibly enriched.

2.6 Application of the Nuclease-Based Screen to *Escherichia coli* and *Streptomyces venezuelae* RNA

When applied to *E. coli* RNA, the method described above also detected 24 non-canonical, unknown species enriched at least 2-fold (Fig. 2.8a). In *S. venezuelae*, this method yielded 28 unknown species that were enriched 2-fold or more (Fig. 2.8b). Independent replicates starting with distinct cell cultures generated enrichment factors with trial-to-trial correlation coefficients of 0.93 in *S. venezuelae* (Fig. 2.8c) and 0.90 in *E. coli* (Fig. 2.8d). None

of the observed unknown species were detected from total *E. coli* or *S. venezuelae* RNA if active nuclease P1 was omitted, or if active nuclease P1 treatment was replaced with incubation in formamide and/or 10 mM EDTA at 95°C, conditions expected to abrogate RNA secondary structure. These results suggest that the species in Fig. 2.8 arise from nuclease-mediated RNA cleavage, and not from the release of small molecules non-covalently associated with RNA.

a

E. coli

observed [M-H] m/z	retention time (min)	nuclease/control enrichment
729.1087	1.7	117
721.1160	5.8	102
668.1103	2.3	100
642.0883	3.7	72
725.1068	1.6	57
643.1063	2.4	55
473.0827	4.0	36
655.1179	4.0	33
820.2175	7.2	23
771.0726	1.4	23
490.1119	5.7	21
789.1588	12.6	20
565.1339	6.5	20
726.0653	4.9	19
760.0740	1.7	17
822.2748	2.7	17
536.0481	1.0	16
824.1958	13.8	14
540.0533	5.3	8
602.1188	6.9	8
965.2397	8.3	7
786.1589	6.8	5
801.1505	8.4	4
654.3013	12.1	4
686.1414	7.1	3
728.1541	7.9	3
811.3491	2.4	2

b

S. venezuelae

observed [M-H] m/z	retention time (min)	nuclease/control enrichment
659.0447	6.8	189
668.1072	2.4	97
602.1252	6.9	51
725.0881	1.1	38
822.2739	2.7	35
565.1754	5.6	28
536.0485	1.0	28
771.0714	1.3	23
616.1401	7.3	22
651.0952	2.0	21
652.0778	4.7	18
372.0320	2.5	14
820.2169	7.2	12
424.0033	1.0	11
750.0549	1.1	11
789.1590	12.6	10
540.0607	5.3	8
859.0962	1.0	8
595.1417	7.0	8
549.0873	5.8	7
572.1117	6.4	6
965.2407	8.2	4
801.1506	8.4	4
686.1436	7.1	4
786.1563	6.6	3
691.1004	2.1	3
654.2997	12.0	3
728.1506	7.9	3
786.1567	6.8	3
816.1701	6.6	2
742.1639	6.8	2
811.3444	2.1	2

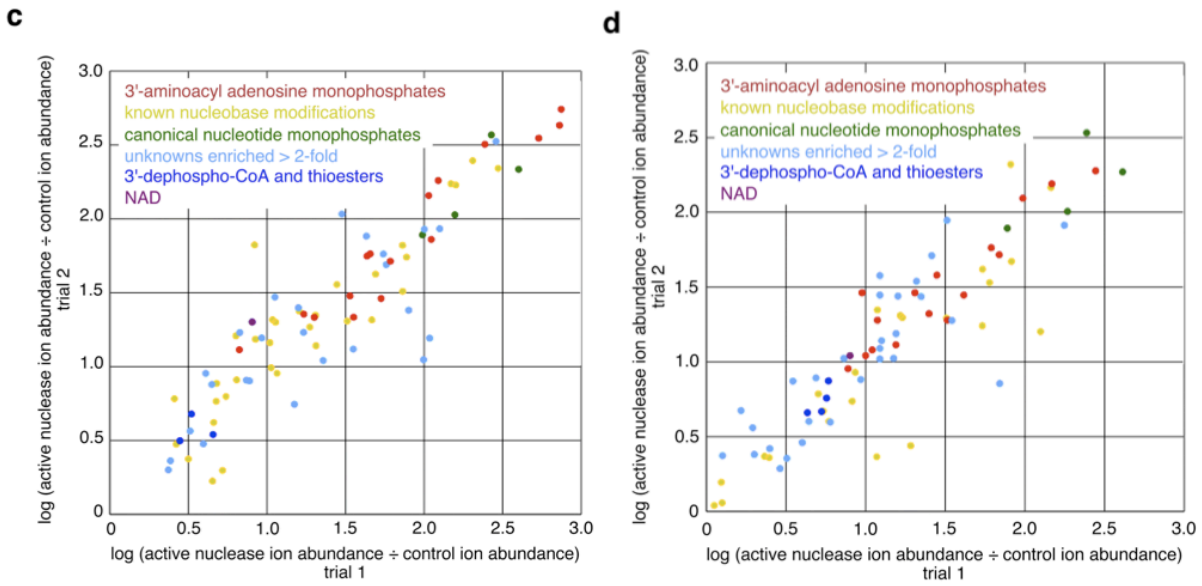


Figure 2.8. (**a**) Unknown species detected by the nuclease digestion method applied to *E. coli* RNA that satisfy the criteria described in the text. (**b**) Unknown species detected by the nuclease digestion method applied to *S. venezuelae* RNA that satisfy the criteria described in the text. The NAD fragment is listed in red, and the CoA derivatives are listed in blue. (**c**) Result of two independent trials (r = 0.90) of the method described in the main text applied to total *E. coli*

Figure 2.8. (continued): RNA. The observed species include 16 3'-aminoacyl adenosine monophosphates, 33 known nucleotide modifications, the four canonical RNA nucleotides, 3'-dephospho-CoA and two thioester derivatives, and 24 additional unknown species with a control:base ratio ≥ 2-fold. (**d**) Results of two independent trials (r = 0.93) of the nuclease method applied to total *S. venezuelae* RNA. The observed species include 16 aminoacyl adenosine monophosphates (red), 25 known nucleobase modifications (yellow), four canonical RNA nucleotides (green), 3'-dephospho-CoA and 3'-dephospho-CoA thioesters (dark blue),[105] NAD (purple), and 24 additional unknown species with a nuclease:control ratio ≥ 2-fold (light blue), 20 of which were not discovered with our previously reported nucleotide cleavage method.[105] In both (**c**) and (**d**) each point represents the average of three experiments.

2.7 Structural Elucidation of *m/z* = 540.0533 as a Fragment of NAD

One unknown species from both *E. coli* and *S. venezuelae* that was enriched 8-fold in the nuclease versus heat-inactivated nuclease samples was $[M-H]^-$ *m/z* = 540.0533 (Fig. 2.9a and b). The data sets generated from culturing *S. venezuelae* in media containing ^{13}C-glucose as the sole carbon source, or in media containing ^{15}N-ammonium sulfate as the sole nitrogen source, resulted in mass increases of this species of 15 Da and 5 Da, respectively (Fig. 2.9c). These results indicated that the unknown species contained 15 carbon atoms and five nitrogen atoms, enabling us to deduce a molecular formula of $C_{15}H_{20}N_5O_{13}P_2$ (expected $[M-H]^-$ *m/z* = 540.0538). The MS/MS spectrum of this ion further indicated that ADP was a major fragment of this ion (Fig. 2.10). We therefore reasoned that the 540.0533 Da species likely consists of a 115.0395 Da group ($C_5H_7O_3$) attached to the pyrophosphate of ADP.

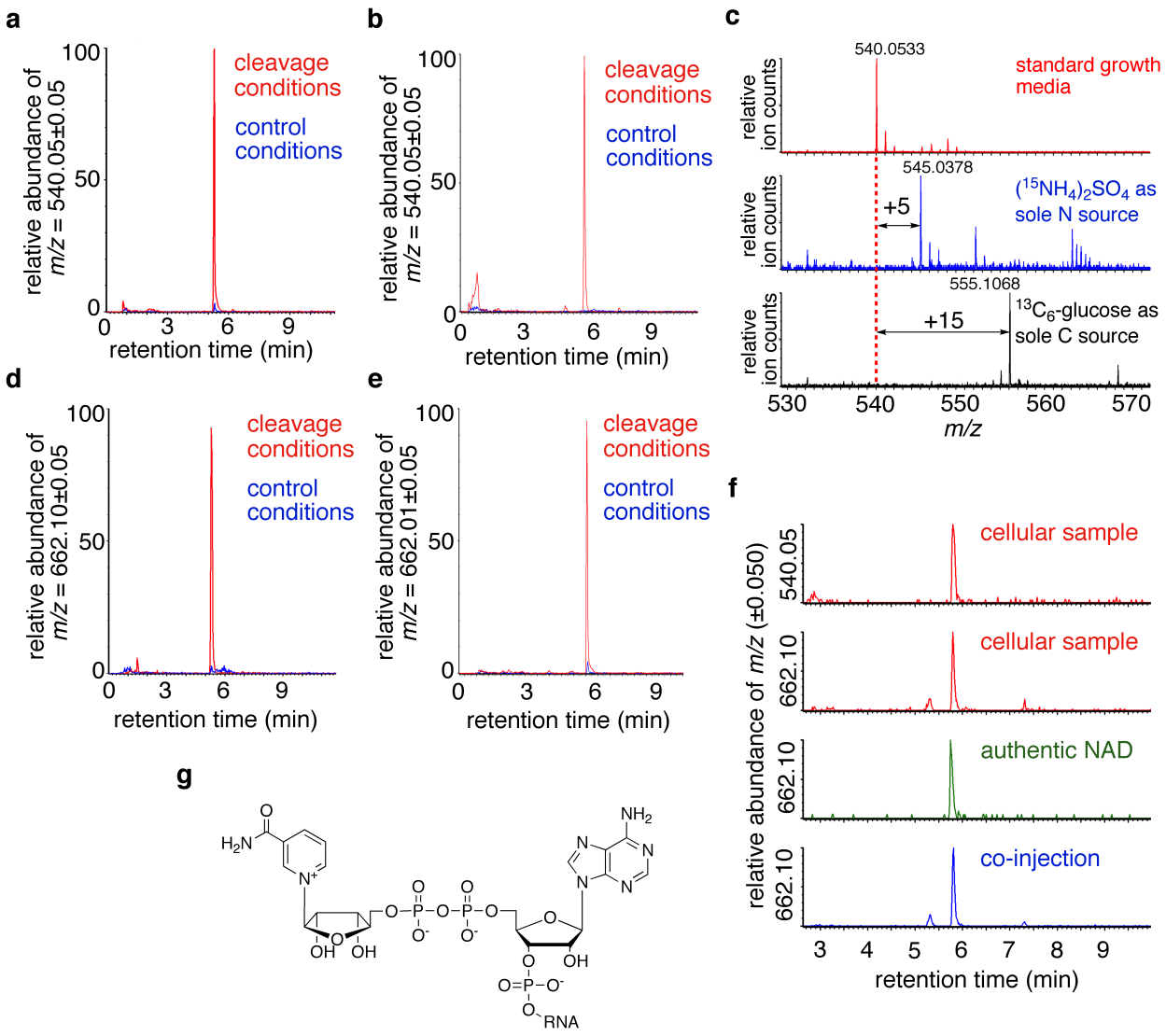


Figure 2.9. A small molecule-linked nucleotide of $[M-H]^-$ m/z = 540.0533 from *E. coli* and *S. venezuelae* RNA. (**a**) The extracted ion chromatograph (EIC) for $[M-H]^-$ m/z = 540.0533 from *E. coli* RNA exposed to active nuclease P1 (cleavage conditions) or to heat-inactivated nuclease P1 (control conditions). (**b**) Same as (a), but using *S. venezuelae* RNA. (**c**) Isotope-labeled *S. venezuelae* total RNA enables determination of the molecular formulas of $[M-H]^-$ m/z = 540.0533. *S. venezuelae* was cultured in media containing ^{13}C-glucose as the sole carbon source, or in media containing ^{15}N-ammonium sulfate as the sole nitrogen source. The resulting RNA species when subjected to the nucleotide cleavage method resulted in mass increases indicating that the $[M-H]^-$ m/z = 540.0533 ion contains 15 carbon atoms and five nitrogen atoms. (**d**) The EICs for $[M-H]^-$ m/z = 662.1018 from *E. coli* RNA or (**e**) from *S. venezuelae* RNA digested with nuclease P1 or control conditions (heat-inactivated nuclease incubation). (**f**) EIC comparison of *E. coli* cellular RNA nuclease P1 digestion products with authentic NAD. (**g**) Plausible structure of NAD-linked cellular RNA. We note that our data is also consistent with RNA attachment to the 2' hydroxyl of NAD, or to either of the nicotinamide-linked ribose hydroxyl groups.

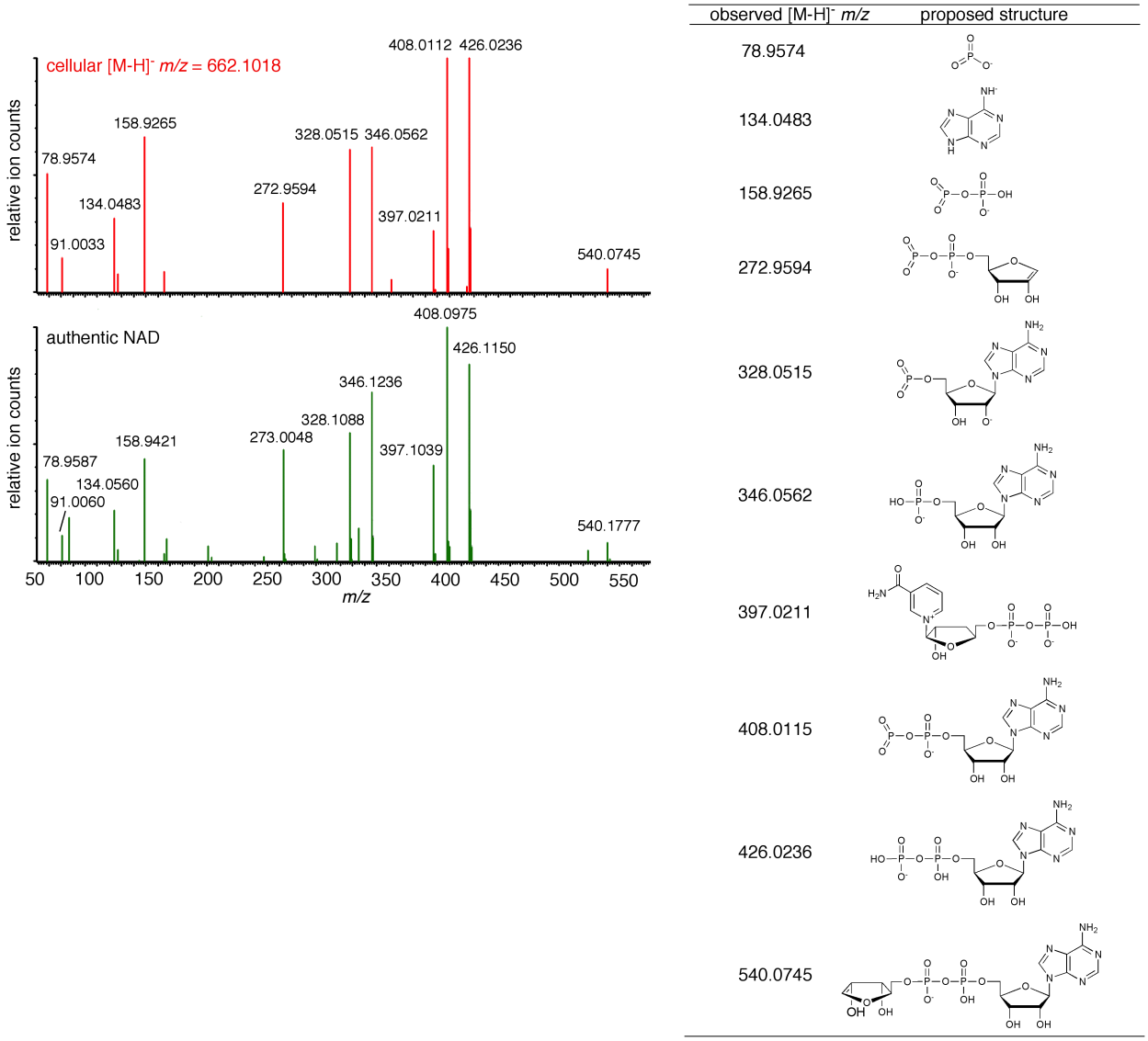


Figure 2.10. MS/MS fragmentation of *E. coli* cellular species $[M-H]^-$ *m/z* = 662.1018 and comparison with authentic NAD confirms the assignment of the $[M-H]^-$ *m/z* = 662.1018 as NAD. A table of plausible fragment ion structures is provided.

The formula $C_5H_7O_3$ suggests that the non-ADP portion of the molecule contains two degrees of unsaturation, consistent with an unsaturated ribose ring. Having considered many possible chemical structures for $C_5H_7O_3$ attached to ADP, we hypothesized that the most plausible source of a five-carbon group conjugated to ADP containing two degrees of unsaturation is an elimination reaction on an adenosine-containing cofactor. We therefore

speculated that the $[M-H]^-$ *m/z* = 540.0533 ion is a breakdown product of a larger small molecule-RNA conjugate that undergoes elimination during MS ionization. In support of this model, an MS analysis performed under milder ionization conditions (Fig. 2.9f) showed a $[M-H]^-$ *m/z* = 662.1032 species that appeared at the same retention time as the $[M-H]^-$ *m/z* = 540.0533 species. As we further adjusted ionization conditions, the signal intensity of $[M-H]^-$ *m/z* = 662.1032 increased to become the predominant peak. Collectively, these observations led us to propose that the $[M-H]^-$ *m/z* = 540.0533 species is a fragment of nicotinamide adenine dinucleotide (NAD) (expected $[M-H]^-$ *m/z* = 662.1018). This hypothesis has been confirmed by LC/MS and MS/MS comparisons of the cellular species with authentic NAD (Fig. 2.9f and Fig. 2.10). Analysis of *E. coli* and *S. venezuelae* RNA using the milder ionization conditions revealed that $[M-H]^-$ *m/z* = 662.1032 was also significantly enriched (8-fold) in the nuclease versus heat-inactivated nuclease samples (Fig. 2.9c and d), in support of these conclusions.

NADH, the reduced form of NAD, is not detected as a small molecule-RNA conjugate by our nuclease-based method. However, the absence of this species (expected $[M-H]^-$ *m/z* = 664.1175) does not rule out the presence of NADH-linked cellular RNA. Indeed, when authentic NADH was added directly to nuclease digestion reactions, a very low NADH signal was observed. Instead, the addition of NADH resulted in an increased NAD signal (Fig. 2.11), suggesting that any NADH-RNA present likely oxidized to NAD-RNA during RNA isolation and sample processing.

	observed NAD (pmol) [M-H]$^-$ *m/z* 662.1018	observed NADH (pmol) [M-H]$^-$ *m/z* 664.1175
authentic NAD (50 pmol) carried through mock RNA isolation	49	0
authentic NADH (50 pmol) carried through mock RNA isolation	45	4
authentic NAD added to nuclease P1-treated total *E. coli* RNA	49	0
authentic NADH added to nuclease P1-treated total *E. coli* RNA	47	2
no authentic NAD or NADH added to nuclease P1-treated total *E. coli* RNA	43	0

Figure 2.11. NADH oxidizes to NAD during RNA isolation, sample processing and analysis. The mock RNA isolation mimicked the total RNA isolation from bacteria protocol, but omitted the isopropanol precipitations and size-exclusion chromatography steps. Observed ion counts are converted into pmol based on NAD and NADH standard curves.

We note that under base cleavage conditions (pH 8) previously used to discover CoA-linked RNA, the abundance of NAD in the nuclease-treated samples is virtually identical (116 vs. 108 ion counts) to the abundance under control conditions (pH 4.5) and thus would have been overlooked using our previous methods.[105]

2.8 NAD is Covalently Linked to RNA

Since NAD is prevalent as a cellular metabolite and cofactor,[130] we sought to confirm that the detected NAD species was not an intracellular contaminant unexpectedly carried through the RNA purification and size exclusion. To ensure that the NAD signal was from a conjugate, we spiked varying quantities of authentic NAD or NADH into *E. coli* and *S. venezuelae* cell lysates, and repeated the RNA isolation, nuclease P1 digestion, and LC/MS analysis. Even though we added up to 10,000-fold more NAD and NADH than was observed in the unspiked samples, the abundance of the corresponding NAD species did not change significantly (Fig. 2.12). These results demonstrate that the NAD species observed in our experiments on *E. coli* and *S. venezuelae* RNA cannot be accounted for by endogenous NAD (or by the oxidation of endogenous NADH), and further support the conclusion that these species arise from cellular small molecule-RNA conjugates.

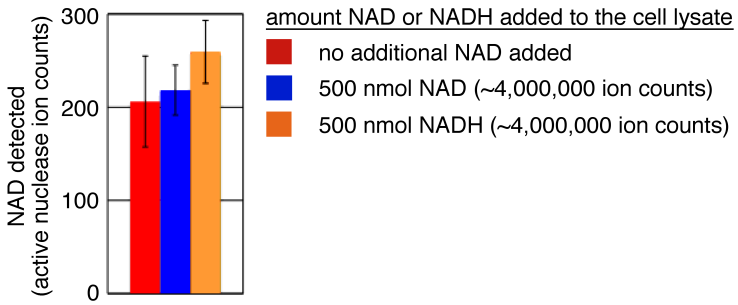


Figure 2.12. Spiking large quantities of NAD or NADH into *E. coli* cell lysate before RNA isolation and treatment with active nuclease P1 does not change the observed ion counts of these species, indicating that the observed NAD signals do not arise from small-molecule NAD contaminants. Error bars represent the standard deviation of three independent trials.

Two nonenzymatic methods to confirm that the NAD signal detected is due not to small molecule contamination but to covalent attachment to NAD, is to subject total RNA to ethanol precipitation and/or size exclusion chromatography. We precipitated total RNA once or twice sequentially, digested with nuclease P1 and measured the NAD signal detected by LC/MS analysis. Since ethanol precipitation removes small molecules from nucleic acids, we expected and indeed observed, stable NAD signal from LC/MS analysis, which indicated NAD was conjugated to RNA (Fig. 2.13).

Modification	*m/z*	Ion Counts After Nuclease Digestion of One Ethanol Precipitation	Ion Counts After Nuclease Digestion of Two Ethanol Precipitation
AMP-Leu	459.13	1840	1250
t6A	491.1	19	13
NAD	540.05	32	10

Figure 2.13. Ion counts of NAD and other RNA modifications remain stable after two ethanol precipitations suggesting that they are indeed covalently linked to RNA. Total RNA was ethanol precipitated, digested with nuclease P1 and followed by LC/MS analysis.

Like ethanol precipitation, size exclusion chromatography separates macromolecules from small molecules. Total RNA was subjected to size exclusion chromatography, where both the macromolecule and small molecule fractions were collected. Each of the fractions were

digested with nuclease P1 and analyzed by LC/MS. If the modifications are covalently linked to RNA, they should only be detected in the macromolecule fraction. Accounting for loss of material through the size exclusion chromatography (as determined by comparing the signal to a known tRNA modification), we observed consistent NAD signal in the appropriate fraction (Fig. 2.14).

Modification	*m/z*	Ion Counts From Macromolecule Fraction	Ion Counts From First Small Molecule Fraction	Ion Counts From Second Small Molecule Fraction
AMP-Leu	459.13	49	n/d above noise	n/d above noise
t6A	491.1	7590	558	19
NAD	540.05	256	n/d above noise	n/d above noise

Figure 2.14. Nucleotide modifications that are covalently linked to RNA are detected in the macromolecule fraction and not the small molecule fraction during size exclusion chromatography. Total RNA was subjected to size exclusion chromatography and the following were collected: macromolecule, first small molecule, and second small molecule fractions. Each fraction was subjected to nuclease P1 digestion and analyzed by LC/MS.

2.9 Characterization of the NAD-RNA Linkage

The structure of NAD led us to hypothesize that NAD is a 5' RNA modification. Because nuclease P1 catalyzes the attack of a water molecule on RNA to generate 5'-phosphonucleotides,[108] all nuclease P1 digestion products other than the nucleotides at the 5' termini undergo a mass shift when the digestion is performed in the presence of isotopically labeled water, compared with digestion in unlabeled water. We carried out nuclease P1 digestion of total RNA from both *E. coli* and *S. venezuelae* in ^{18}O-enriched water. As expected, 3'-Phe-AMP exhibited a +2 Da mass shift when the digestion was carried out in ^{18}O water compared to in ^{16}O water (observed $[M-H]^-$ *m/z* = 495.1286; expected $[M-H]^-$ *m/z* = 495.1285) (Fig. 2.15a). In contrast, there was no mass shift for NAD in both isotopically-labeled and unlabeled water (Fig. 2.15b). This finding is consistent with a model in which NAD is present at the 5' termini of RNA (Fig. 2.9g).

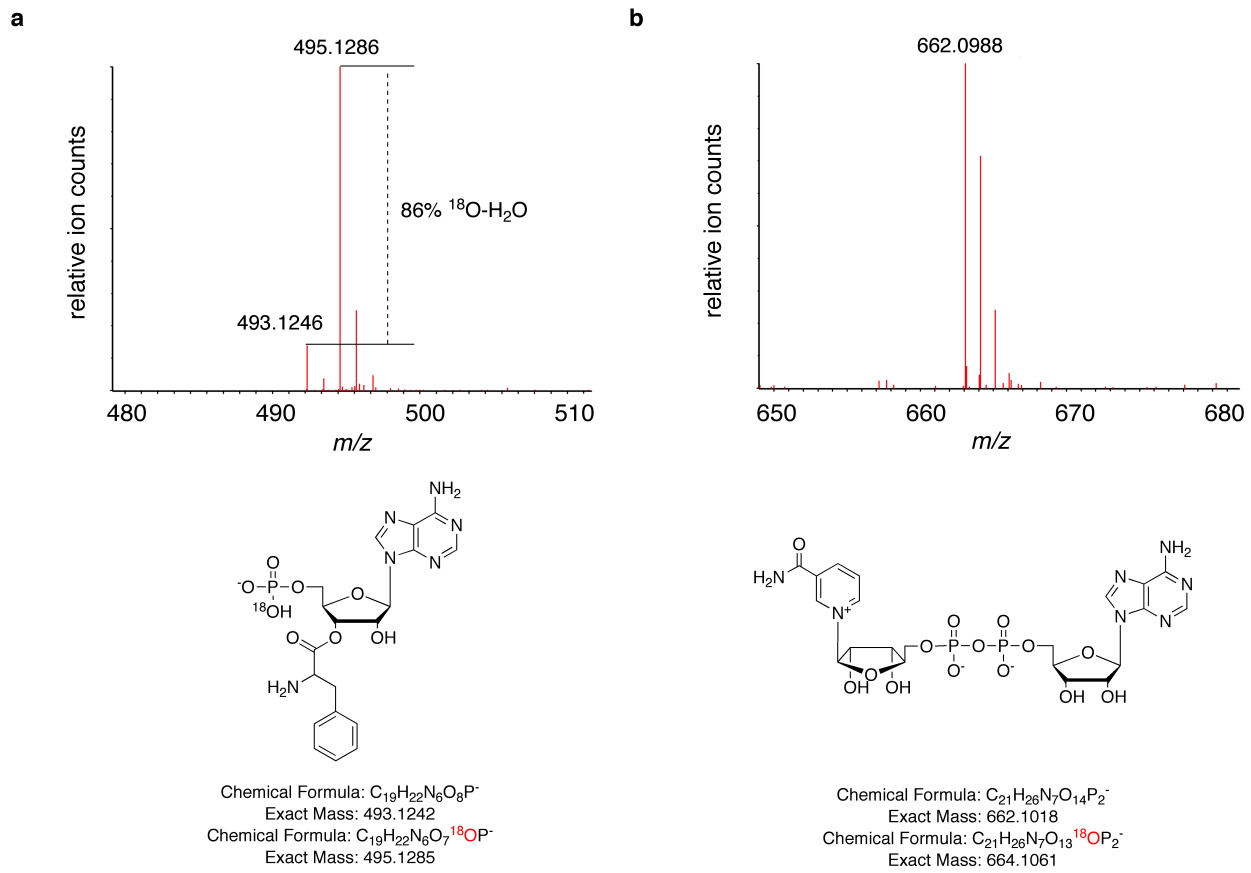


Figure 2.15. In the presence of ^{18}O water, all nuclease P1 digestion products other than the nucleotides at the 5' ends of substrates will exhibit a +2 Da mass shift compared with the products arising from digestion in ^{16}O water. (**a**) Phenylalanyl AMP, a known small molecule-RNA conjugate that exists at the 3' end of tRNA, exhibits the expected +2 Da shift. (**b**) No mass shift is observed in the case of NAD, indicating that this modification is located at the 5' terminus of the RNA(s).

Unlike nuclease P1, RNase A mediates RNA cleavage by catalyzing the attack of a 2'-hydroxyl group to form a 2', 3'-cyclic phosphonucleotide, which then can be hydrolyzed to generate a nucleotide 3'-monophosphate. Products of RNase A digestion, with the exception of the 3'-terminal base, are therefore monophosphonucleotides.[131-133] When total RNA from *E. coli* and *S. venezuelae* was digested with RNase A, the internal rRNA nucleoside modification $N^6,N^{6'}$-dimethyladenine[134] was detected as a mixture of the cyclic and acyclic monophosphonucleotides, as expected (observed $[M-H]^-$ m/z = 356.0772 and m/z = 374.0873,

expected $[M-H]^-$ *m/z* = 356.0760, and *m/z* = 374.0866). RNase A digestion also generated monophosphorylated NAD (observed $[M-H]^-$ *m/z* = 742.0678, expected $[M-H]^-$ *m/z* = 742.0682). Similar results were observed when authentic NAD-linked RNA, generated by T7 RNA polymerase[135] was digested with RNase A (observed $[M-H]^-$ *m/z* = 742.0653). These results are consistent with the nuclease P1 digestion results and further support a model in which the NAD group is covalently linked not to the 3' terminus but to the 5' terminus of RNA.

2.10 Cellular NAD-RNA Is Surprisingly Abundant

We quantified the amount of NAD-linked RNA in *E. coli* cells using two methods. First, we generated a standard curve that relates known quantities of authentic NAD (quantified by spectrophotometry) to observed ion counts under both the original and the milder ionization conditions. Ion counts from *E. coli* RNA when plotted on the resulting curve result in the estimate of ~3,000 copies of NAD-RNA per *E. coli* cell for both ionization conditions (Fig. 2.16a and b). In a second approach, we added a known quantity of authentic NAD to the biological nuclease P1-digested RNA sample before LC/MS analysis, and used the resulting increase in the NAD ion count to relate added NAD concentration and observed signal. This second method resulted in a similar estimate of ~3,300 copies of NAD-RNA per *E. coli* cell (Fig. 2.16b). The estimated abundance level is comparable to that of aminoacylated tRNAs in the second highest quartile of abundance in *E. coli* such as Ile-charged $tRNA^{Ile}$ (~2,800 copies per cell), Ala-charged $tRNA^{Ala}$ (~3,100 copies per cell), or Thr-charged $tRNA^{Thr}$ (~2,100 copies per cell) (Fig. 2.17).[136] Therefore, these results suggested that NAD-linked RNA is one of the more abundant known cellular small molecule-RNA conjugates discovered to date. We observed 87

fmol of NAD-RNA per μg of *E. coli* RNA and 142 fmol of NAD-RNA per μg of *S. venezuelae* RNA.

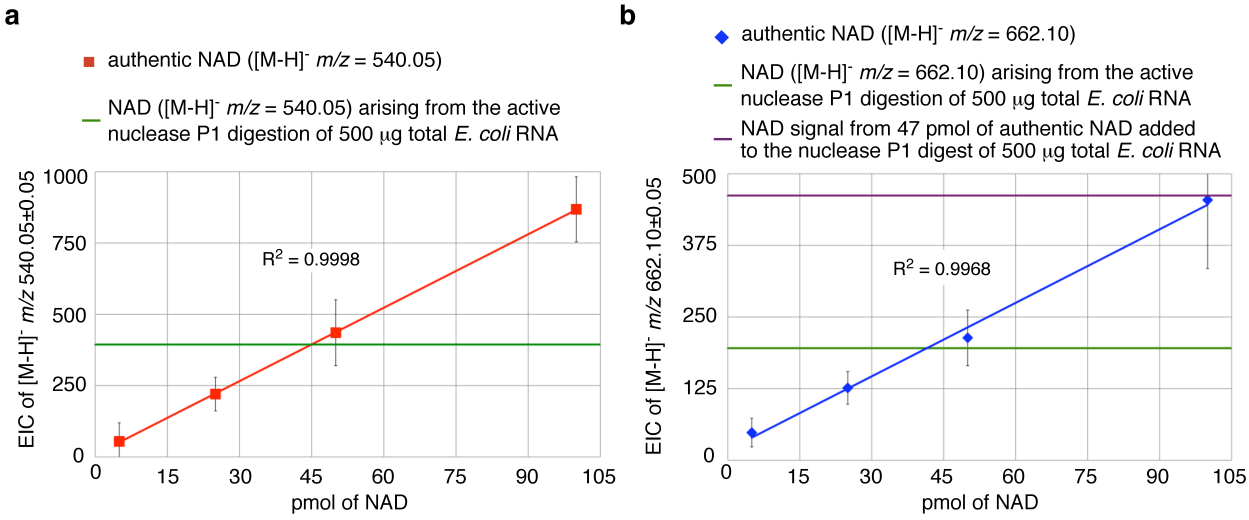


Figure 2.16. Standard curve relating known NAD quantities to the abundance of the major daughter ion ([M-H]⁻ *m/z* = 540.0538) or of the parent ion ([M-H]⁻ *m/z* = 662.1018). These curves were used to determine the amount of NAD-RNA per *E. coli* cell. The amount of NAD signal arising from nuclease P1 digestion of total *E. coli* RNA is indicated with the green lines. The purple line indicates the observed NAD signal when 47 pmols of authentic NAD is added to nuclease P1-digested total *E. coli* RNA. All three cases indicate that there are ~43-47 pmol of NAD-RNA per 500 μg of *E. coli* RNA, which corresponds to ~3,000-3,300 copies of NAD-RNA per *E. coli* cell. Error bars in **a** and **b** represent the standard deviation of three independent trials.

charged tRNA	molecules/cell
Leu	9023
Arg	6678
Gly	6496
Val	5105
Glu	4717
Ala	3867
Ser	3812
Ile	3474
Thr	2656
Met	2632
Asp	2396
Pro	2201
Tyr	2030
Lys	1924
Gln	1645
Cys	1587
Asn	1193
Phe	1037
Trp	943
His	639
tRNA (total)	205,000
tmRNA	700
23S rRNA	18,700
16S rRNA	18,700
5S rRNA	18,700
mRNA	4,000
RNA from primase	11

Figure 2.17. Abundance of different RNAs in an *E. coli* cell, adapted from Jakubowski and Goldman (1984). *J Bacteriol* 158(3): 769-776.[136]

2.11 Transcriptional Initiation by *E. coli* RNA Polymerase *In Vitro* Cannot Account for Observed Levels of NAD-RNA

Structural similarities between NAD and ATP led us to speculate that NAD might be incorporated into RNA at the 5'-terminus through aberrant transcriptional initiation with NAD or its reduced form instead of ATP. Indeed, NAD has been incorporated into the 5'-terminus of RNA transcripts *in vitro* using T7 RNA polymerase.[135] To test if transcriptional initiation was responsible for incorporating NAD into RNA transcripts, we used *E. coli* RNA polymerase to carry out *in vitro* transcription in the presence of high concentrations of NADH using two

templates. The first template was a modified pUC19 plasmid, which encoded an adenosine at the +1 position of each of its four predicted transcripts. An *in vitro* transcription reaction containing 0.5 mM of each NTP and either 0.5 mM or 5 mM of NADH yielded 222 μg or 174 μg of RNA, respectively. When this RNA was purified, digested with nuclease P1, and analyzed by LC/MS, no NAD or NADH was detected (Fig. 2.18). The second template used was *E. coli* genomic DNA. *In vitro* transcription in the presence of either 0.5 mM or 5 mM of NADH yielded 78 μg or 70 μg of RNA. Once again, this material contained no detectable NAD or NADH after nuclease P1 digestion (Fig. 2.18). In contrast, when an authentic 5'-NAD-linked transcript (generated using T7 RNA polymerase) was spiked into an *in vitro* transcription reaction and processed in the same way, NAD-linked RNA was readily detected (Fig. 2.18).

DNA template	spiked NAD-RNA (pmol)	*E. coli* RNA pol.	NADH (mM)	RNA yield (µg)	expected ion counts of NAD or NADH from transcription	expected ion counts of NAD or NADH from spiked NAD-RNA	observed ion counts of NAD or NADH
pUC19+1A	0	-	0.5	1.5	0	0	8
pUC19+1A	0	+	0.5	222.4	≥ 636	0	6
pUC19+1A	0	+	5.0	173.8	≥ 495	0	11
E. coli genomic DNA	0	-	0.5	1.7	0	0	8
E. coli genomic DNA	0	+	0.5	78.4	≥ 217	0	6
E. coli genomic DNA	0	+	5.0	69.8	≥ 192	0	3
E. coli genomic DNA	1.0	-	0.5	2.2	0	9	8
E. coli genomic DNA	1.0	+	0.5	86.4	≥ 240	9	10
E. coli genomic DNA	2.3	-	0.5	2.0	0	20	16
E. coli genomic DNA	2.3	+	0.5	72.9	≥ 201	20	22
E. coli genomic DNA	4.5	-	0.5	1.6	0	38	45
E. coli genomic DNA	4.5	+	0.5	91.0	≥ 254	38	48

Figure 2.18. *In vitro* transcription with *E. coli* RNA polymerase in the presence of NADH does not generate significant quantities of NAD-linked RNA. *E. coli* RNA polymerase was used to generate RNA *in vitro* from plasmid or genomic DNA templates in the presence of 0.5 mM or 5.0 mM of NADH. When the resulting RNA was digested with nuclease P1 and analyzed by LC/MS, the observed (NAD+NADH) signal (column 8) was much lower than the expected ion counts if the observed quantities of NAD-RNA in *E. coli* arise from aberrant transcriptional initiation with NAD or NADH (column 6). Spiking 1.0, 2.3, or 4.5 pmol of authentic NAD-linked RNA (generated with T7 RNA polymerase; see the main text) into the *in vitro* transcription reaction resulted in observed ion counts that were consistent with, but not significantly greater than, the expected levels of NAD ions from the spiked material (column 7). Column 6 lists the expected ion counts of (NAD+NADH) assuming that the fraction of RNA generated from *in vitro* transcription under these conditions is at least as great as that observed from *E. coli* cellular RNA.

Based on the above estimated abundances of NAD-RNA in *E. coli*, we expect to obtain more than 19 pmol of NAD from ~222 µg of A-initiated RNA, and 6.8 pmol of NAD from ~78 µg of RNA transcribed from the *E. coli* genome, if transcriptional initiation were predominantly responsible for the NAD-RNA conjugates. Such quantities would be readily detected by our methods, which can reliably detect less than or equal to 0.5 pmol of NAD. If the inability of *E. coli* RNA polymerase to incorporate these levels of NAD *in vitro* indicates the inability to do so in cells, NAD groups are likely installed following transcriptional initiation.

2.12 NAD Radical Formation May Contribute to the Unexpectedly High +1 Da Isotope Peak in the Authentic and Cellular NAD Spectra

Based on the natural isotope abundances, the expected abundance ratio of the +1 Da isotope peak at $[M-H]^-$ *m/z* = 663.1069 to the isotope peak at $[M-H]^-$ *m/z* = 662.0986 is equal to 26.8%. However, the observed ratio is higher for both authentic and cellular RNA-derived NAD samples (Fig. 2.19a). The increase in the $[M-H]^-$ *m/z* = 663.1069 signal might be due to the contributions of other species that either coelute with NAD or are produced during ionization. For example, an NAD radical anion (Fig. 2.19b) would be observed primarily as a $[M-H]^-$ *m/z* = 663.1091 species. A mixture of NAD anion and NAD radical anion species during ionization would result in a higher-than-expected abundance of the NAD +1 Da peak.

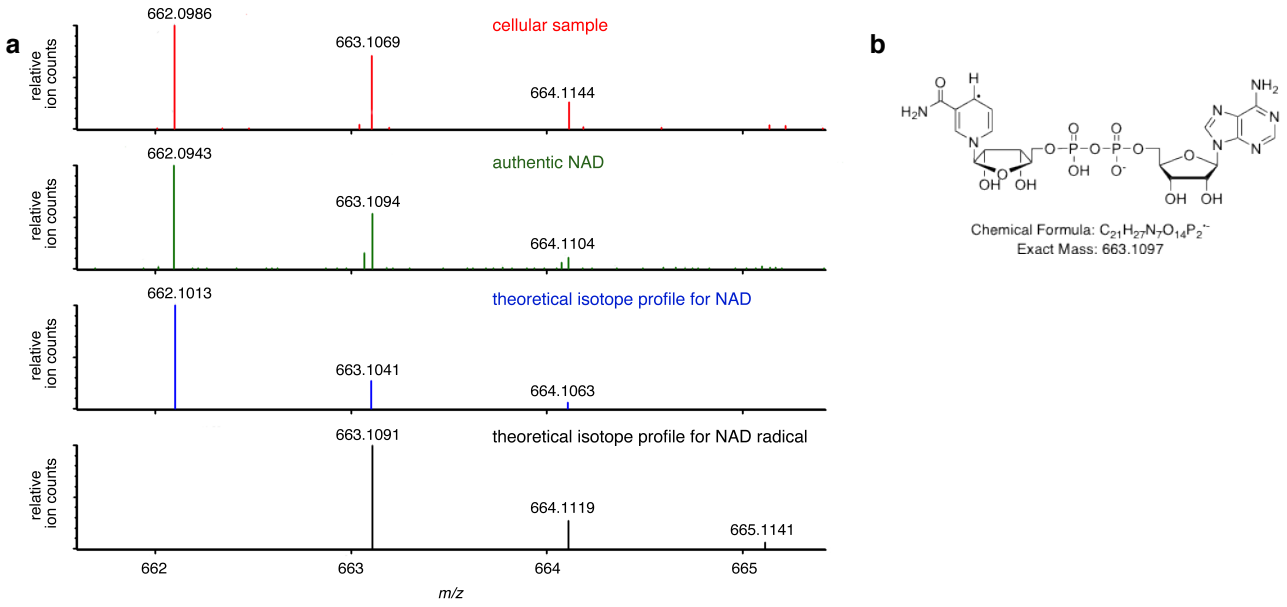


Figure 2.19. NAD radical may contribute to the $[M-H]^-$ *m/z* 663.1069 signal in the authentic and cellular NAD spectra. (**a**) Mass spectra of cellular and authentic NAD, and the theoretical isotope profiles for NAD and NAD radical. (**b**) Structure of NAD radical.

2.13 Conclusions

The method developed in this work enables the detection, in principle, of any small molecule-RNA conjugate. The application of this method resulted in the discovery of NAD-RNA in *E. coli* and *S. venezuelae*, bacteria from two different phyla, as well as in the detection of a variety of additional unknown, non-canonical small molecule-RNA conjugates. Following the recent discovery of CoA-linked RNA, these results further indicate that the chemical diversity of biological RNA is even greater than suggested by such known examples as aminoacylated tRNAs, RNAs containing modified nucleobases, and, in eukaryotes, 5'-capped mRNA.

We note that under base cleavage conditions (pH 8) previously used to discover CoA-linked RNA, the abundance of NAD in the nuclease-treated samples is virtually identical (116 vs. 108 ion counts) to the abundance under control conditions (pH 4.5) and thus would have been overlooked using our previous approaches.[105] The inability to detect NAD-linked RNA under base cleavage conditions necessitates a more general method for the discovery of less reactive small molecule-RNA conjugates as developed in this work.

The NAD group is linked to RNA(s) at the 5' terminus. On average we observe ~3,000 NAD-RNA molecules per *E. coli* cell, which suggests that NAD-linked RNAs together are approximately as abundant as Phe-linked tRNA in *E. coli*,[136] ~4-fold more abundant than *E. coli* tmRNA,[137] and ~300-fold more abundant than the short RNAs generated by DNA primase during DNA synthesis.[138,139] Our *in vitro* transcription experiments suggest that non-specific transcriptional initiation is not the primary mechanism for NAD-RNA formation. It is possible that additional cellular components are necessary for a non-specific transcription initiation pathway for NAD incorporation. Our results also do not rule out a gene-specific transcription initiation pathway for NAD. Studies are ongoing to identify additional small molecule-RNA

conjugates, to characterize the RNA species to which these groups are attached, and to evaluate functional roles of such groups in the cell.

2.14 Experimental Methods

General. Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification.

Bacterial Growth and Crude Nucleic Acid Isolation. *E. coli* TOP10 (Invitrogen) was cultured to OD_{600} = 0.7-0.8 at 37ºC in 2 L LB broth Miller (EMD Bioscience). *S. venezuelae* ATCC #10595 was cultured to OD_{600} = 0.7-0.8 at 30 ºC in 2 L MYME media (100 g/L sucrose, 10 g/L maltose, 5 g/L peptone, 3 g/L yeast extract, 3 g/L malt extract).[140] All subsequent manipulations of cells were carried out on ice or at 4 °C. The bacteria were centrifuged (10 min at 6,750 x g), resuspended in lysis buffer (1% SDS, 2 mM EDTA, 32 mM NaOAc, pH 4.5), and vortexed vigorously. After incubation on ice for 15 min, the lysate was cleared by centrifugation (10 min at 10,000 x g), extracted with acid-phenol chloroform (Ambion) until the organic-aqueous interface was clear, and the aqueous layer was washed once with chloroform. An equal volume of isopropanol was added to the resulting aqueous extract and the mixture was incubated on crushed dry ice 20-30 min prior to centrifugation (20 min, 15,000 x g). The resulting pellet was dissolved in 50 mM NH_4OAc, pH 4.5, and subjected to size-exclusion chromatography using NAP5 columns (GE Healthcare). The macromolecular fraction was treated with 0.04 U/µL TURBO DNase (Ambion) at 25 ºC for 30 min and then with 0.06 U/µL proteinase K (New England Biolabs) at 25 °C for 30 min. The resulting solution was extracted with acid-phenol chloroform (Ambion) twice, washed with chloroform, and again subjected to size-exclusion

chromatography using NAP5 columns. The macromolecular fraction was divided into aliquots, lyophilized, and stored as a dry powder at -80 ºC.

Nuclease Digestion. The cleavage condition sample was prepared by incubating 500 µg of *E. coli* RNA or 500 µg of *S. venezuelae* RNA with 10 U nuclease P1 (Sigma-Aldrich) in 200 µL of 50 mM NH_4OAc, pH 4.5 at 37 ºC for 40 min. The control condition samples was prepared by incubating 500 µg aliquot of *E. coli* RNA or *S. venezuelae* RNA with 10 U heat-inactivated nuclease P1 (95 ºC for one hour) in 200 µL of 50 mM NH_4OAc, pH 4.5 at 37 ºC for 40 min. The digestion products were purified by size-exclusion chromatography (NAP5) and the small-molecule fraction was retained. The nuclease P1 digestion with $H_2{}^{18}O$ (Cambridge Isotope Laboratories) was performed as described above except in buffer with a final composition containing 86% $H_2{}^{18}O$ and 14% $H_2{}^{16}O$.

LC/MS Data Collection and Analysis. LC/MS was performed using a Waters Aquity UPLC Q-TOF Premier instrument with an Aquity UPLC BEH C18 column (1.7 µm, 2.1 mm x 100 mm, Waters). Mobile phase A was 0.1% aqueous ammonium formate, and mobile phase B was 100% methanol. The flow rate was a constant 0.300 mL/min and the mobile phase composition was as follows: 0% B for 3 min; linear increase over 17 min to 100% B; maintain at 100% B for 2 min before returning linearly to 0 % B over 1 min. Electrospray ionization (ESI) was used, with a capillary voltage of 3.5 kV, sampling cone voltage of 40.0 kV, and collision voltage of 1.0 eV. The drying gas temperature was 300 ºC, the drying gas flow rate was 800 L/hour, the source temperature was 150 ºC, and the detector was operated in negative ion mode. To observe the NAD parent ion ($[M-H]^-$ m/z = 662.1018), the capillary voltage was 2.75 kV and the sampling cone voltage was 20.0 kV. For each sample, 15 µL of the redissolved lyophilized material was

injected. Ions with cleavage condition average integrated ion intensities below 50 ion counts were not considered for further analysis. The XCMS program[109] quantified the area under detected ion abundance peaks as it stepped through each ion chromatogram; the step size was set to 0.050 Da. Integrated ion abundances were averaged among replicates, and the enrichment values reported were the ratios of these average ion intensities between active nuclease conditions and heat-inactivated nuclease (control) conditions.

MS/MS Fragmentation Analysis. MS/MS experiments were performed using the same instrument, LC gradient, and MS parameters described above. The collision voltage was varied empirically from 20.0-30.0 2V.

Isotope Labeling of *S. venezuelae* RNA. *S. venezuelae* ATCC #10595 was cultured to OD_{600} = 0.7-0.8 at 30 °C in minimal media (4.3 g/L NaH_2PO_4, 6.1 g/L K_2HPO_4, 2.0 g/L NaCl, 2.0 mg/L $FeSO_4 \cdot 7H_2O$, 2.0 mg/L $MnCl_2 \cdot 4H_2O$, 2.0 mg/L $ZnSO_4 \cdot 7H_2O$, 2.0 mg/L $CaCl_2$, 0.6 g/L $MgSO_4 \cdot 7H_2O$, 10.0 g/L glucose, and 2.0 g/L $(NH_4)_2SO_4$)[140]. For ^{13}C labeling, ^{13}C-glucose (99% ^{13}C, Cambridge Isotope Laboratories) was used as the sole carbon source. For ^{15}N labeling, ^{15}N-ammonium sulfate (99% ^{15}N, Cambridge Isotope Laboratories) was used as the sole nitrogen source. RNA isolation, nuclease P1 digestion, and LC/MS analysis was performed as described above.

NAD Spiking Into Cell Lysate. *E. coli* TOP10 cell lysate was prepared as described above and divided into three equal aliquots. NAD and NADH (Sigma-Aldrich) were added to the cell lysate as follows: aliquot 1: no added NAD or NADH; aliquot 2: 500 nmol of NAD; aliquot 3:

500 nmol of NADH. Based on our standard curve analysis 500 nmol of NADH or 500 nmol of NAD theoretically represents 4,000,000 ion counts; in practice, this amount of NAD or NADH saturates the detector. Cell lysates were processed and analyzed as described above.

RNase A Digestion. Total RNA prepared as described above was subjected to digestion with 1 μg RNase A (Ambion) in 200 μL of 50 mM NH_4OAc, pH 4.5, at 37 ºC for 20 min). After digestion, the sample was subjected to size-exclusion chromatography (NAP5, GE Healthcare) and the small-molecule fraction was lyophilized. The lyophilized product was redissolved in 20 μL of 0.1 % aqueous ammonium formate and analyzed by LC/MS as described above.

Number of NAD-RNA(s) per *E. coli* cell. To establish a standard curve of MS ion counts per pmol NAD, 5, 25, 50 and 100 pmols of authentic NAD (Sigma-Aldrich) were analyzed by initial and mild ionization conditions as described above. The NAD signal from 500 μg of total *E. coli* RNA was compared to the standard curves to determine the number of pmols of cellular NAD per μg RNA. The molecules of NAD-RNA per cell was then calculated based on $59x10^{-15}$ g RNA per *E. coli* cell.[141]

Authentic NAD and NADH Carried Through Mock RNA Isolation. To determine if cellular NADH-RNA could be detected by our nuclease based method, 50 pmol of NAD and NADH (Sigma-Aldrich) in 500 μL of 50 mM NH_4OAc, pH 4.5 were incubated on ice for 15 min, and then at 4 °C for 10 min. The samples were moved back to ice for 5 min and then placed at 4 °C for 10 min. This procedure was repeated three times, corresponding to the number of acid-phenol chloroform and chloroform extractions during the crude nucleic acid isolation. The NAD

and NADH samples were then placed on crushed dry ice for 20-30 min prior to incubation at 4 °C for 20 min, mimicking the isopropanol precipitation of cellular RNA. While the crude nucleic acids were subjected to size-exclusion chromatography, the NAD and NADH samples were exposed to air and incubated at 25 °C for 10-15 min. The samples were then incubated at 25 °C for 1 h, and lyophilized. The dry powder was redissolved in 200 μL of 50 mM NH_4OAc, pH 4.5 and incubated at 37 ºC for 40 min. The NAD and NADH samples were exposed to air and incubated at 25 ºC for 15-20 min. The samples were lyophilized and stored as a dry powder at -20 ºC.

Preparation of Authentic 5'-NAD-RNA. 5'-NAD-RNA was prepared as previously described.[135] A DNA template containing the T7 class II promoter (Φ2.5) and encoding a 228-base RNA was prepared by PCR. Transcription was carried out in 1x New England Biolabs (NEB) RNA pol buffer supplemented with 1 mM each NTP, 1 mM NADH (Aldrich), 0.01% Triton X-100, 5 mM DTT, 0.2 U/μl RNase inhibitor (NEB), 0.2 μM DNA template, and 5 U/μL T7 RNA polymerase (NEB). The reactions were incubated at 37 °C for 2 hrs before the product was isolated by precipitation with ethanol, dissolved in TURBO-DNase buffer (Ambion), and treated with TURBO-DNase (0.04 U/μL final concentration, 37 °C, 30 min). The DNase-treated RNA was precipitated with ethanol, redissolved in water, and purified by size-exclusion chromatography (NAP5, GE Healthcare). The precipitated NAD-RNA product following DNase treatment was dissolved in water, purified by silica column (Qiagen RNeasy), quantified by A_{260}, digested with nuclease P1, and analyzed by LC/MS as described above in order to estimate the fraction of the resulting RNA strands that contained the NAD modification. We observed that ~3% of the transcripts generated by this procedure were linked to NAD.

***In Vitro* Transcription.** *In vitro* transcription reactions contained *E. coli* RNA polymerase (Epicentre Biotechnologies) in 50 mM Tris-HCl, pH 7.5; 150 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100 supplemented with NTPs (0.5 mM each), DTT (10 mM), and NADH (0.5 mM or 5.0 mM). The plasmid or genomic DNA template was added to a final concentration of 0.02 μg/μL and *E. coli* RNA polymerase was added to a final concentration of 0.04 U/μL. The reactions were incubated at 37 °C for 22 hrs before the product was isolated by precipitation with ethanol, dissolved in TURBO-DNase buffer (Ambion), and treated with TURBO-DNase (Ambion) at a final concentration of 0.04 U/μL for 30 minutes at 37 °C. The RNA was purified by RNeasy spin-column (Qiagen) to remove free NTPs and NADH. The resulting RNA was digested with nuclease P1 and analyzed by LC/MS as described above.

The first template was a modified pUC19 plasmid, which encoded an adenosine at the +1 position of each of its four predicted transcripts. An *in vitro* transcription reaction containing 0.5 mM of each NTP and either 0.5 mM or 5 mM of NADH yielded 222 μg or 174 μg of RNA, respectively. The second template used was *E. coli* genomic DNA. *In vitro* transcription in the presence of either 0.5 mM or 5 mM of NADH yielded 78 μg or 70 μg of RNA. Once again, this material contained no detectable NAD or NADH after nuclease P1 digestion.

2.15 References

1 Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A* 106, 7768-7773, doi:0900528106 [pii]10.1073/pnas.0900528106 (2009).

2 Crain, P. F. & James, A. M. in *Methods in Enzymology* Vol. Volume 193 782-790 (Academic Press, 1990).

3 Crain, P. F. Mass spectrometric techniques in nucleic acid research. *Mass Spectrometry Reviews* 9, 505-554 (1990).

4 Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins* 32, 414-424, doi:10.1002/(SICI)1097-0134(19980901)32:4<414::AID-PROT2>3.0.CO;2-G [pii] (1998).
5 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78, 779-787, doi:10.1021/ac051437y (2006).
6 Woese, C. R. Bacterial evolution. *Microbiol Rev* 51, 221-271 (1987).
7 Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* 47, 621-669 (1983).
8 Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87, 4576-4579 (1990).
9 Limbach, P. A., Crain, P. F. & McCloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res* 22, 2183-2196 (1994).
10 Edmonds, C. G. *et al.* Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). *J Bacteriol* 173, 3138-3148 (1991).
11 Dunin-Horkawicz, S. *et al.* MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* 34, D145-149, doi:34/suppl_1/D145 [pii]10.1093/nar/gkj084 (2006).
12 Ng, W. V. *et al.* Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci U S A* 97, 12176-12181, doi:10.1073/pnas.190337797 [pii] (2000).
13 Deamer, D. W. *Light transducing membranes, structure, function, and evolution*. (Academic Press, 1978).
14 I.E.D, D. in *Advances in Microbial Physiology* Vol. Volume 15 eds A. H. Rose & D. W. Tempest) 85-120 (Academic Press, 1977).
15 Soppa, J. From genomes to function: haloarchaea as model organisms. *Microbiology* 152, 585-590, doi:152/3/585 [pii]10.1099/mic.0.28504-0 (2006).
16 Wang, G., Kennedy, S. P., Fasiludeen, S., Rensing, C. & DasSarma, S. Arsenic resistance in Halobacterium sp. strain NRC-1 examined by using an improved gene knockout system. *J Bacteriol* 186, 3187-3194 (2004).
17 Peck, R. F., DasSarma, S. & Krebs, M. P. Homologous gene knockout in the archaeon Halobacterium salinarum with ura3 as a counterselectable marker. *Mol Microbiol* 35, 667-676, doi:1739 [pii] (2000).
18 Zaigler, A., Schuster, S. C. & Soppa, J. Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon Haloferax volcanii. *Mol Microbiol* 48, 1089-1105, doi:3497 [pii] (2003).
19 Muller, J. A. & DasSarma, S. Genomic analysis of anaerobic respiration in the archaeon Halobacterium sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *J Bacteriol* 187, 1659-1667, doi:187/5/1659 [pii]10.1128/JB.187.5.1659-1667.2005 (2005).
20 Klein, C. *et al.* The membrane proteome of Halobacterium salinarum. *Proteomics* 5, 180-197, doi:10.1002/pmic.200400943 (2005).
21 Robb, F. T., Place, A. R., DasSarma, S. & Fleischmann, E. M. *Archaea: a laboratory manual. Halophiles*. (Cold Spring Harbor Laboratory, 1995).
22 Wood, V. *et al.* The genome sequence of Schizosaccharomyces pombe. *Nature* 415, 871-880 (2002).

23 Christie, K. R. *et al.* Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res* 32, D311-314, doi:10.1093/nar/gkh03332/suppl_1/D311 [pii] (2004).
24 Mewes, H. W. *et al.* Overview of the yeast genome. *Nature* 387, 7-65, doi:10.1038/42755 (1997).
25 Goffeau, A. *et al.* Life with 6000 genes. *Science* 274, 546, 563-547 (1996).
26 Bokar, J. in *Fine-Tuning of RNA Functions by Modification and Editing* 141-177 (2005).
27 Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* 24, 1558-1564, doi:nbt1268 [pii]10.1038/nbt1268 (2006).
28 Pollak, N., Dolle, C. & Ziegler, M. The power to reduce: pyridine nucleotides--small molecules with a multitude of functions. *Biochem J* 402, 205-218, doi:BJ20061638 [pii]10.1042/BJ20061638 (2007).
29 Roberts, G. C., Dennis, E. A., Meadows, D. H., Cohen, J. S. & Jardetzky, O. The mechanism of action of ribonuclease. *Proc Natl Acad Sci U S A* 62, 1151-1158 (1969).
30 Usher, D. A. On the mechanism of ribonuclease action. *Proc Natl Acad Sci U S A* 62, 661-667 (1969).
31 Usher, D. A., Richardson, D. I., Jr. & Eckstein, F. Absolute stereochemistry of the second step of ribonuclease action. *Nature* 228, 663-665 (1970).
32 Grosjean, H. B. R. *Modification and editing of RNA*. (ASM Press, 1998).
33 Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res* 31, e8 (2003).
34 Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J Bacteriol* 158, 769-776 (1984).
35 Moore, S. D. & Sauer, R. T. Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli. *Mol Microbiol* 58, 456-466, doi:MMI4832 [pii]10.1111/j.1365-2958.2005.04832.x (2005).
36 Frick, D. N. & Richardson, C. C. DNA primases. *Annu Rev Biochem* 70, 39-80, doi:70/1/39 [pii]10.1146/annurev.biochem.70.1.39 (2001).
37 Ogawa, T., Hirose, S., Okazaki, T. & Okazaki, R. Mechanism of DNA chain growth XVI. Analyses of RNA-linked DNA pieces in Escherichia coli with polynucleotide kinase. *J Mol Biol* 112, 121-140 (1977).
38 Tsukiji, S., Pattnaik, S. B. & Suga, H. Reduction of an aldehyde by a NADH/Zn2+ -dependent redox active ribozyme. *J Am Chem Soc* 126, 5044-5045, doi:10.1021/ja0495213 (2004).
39 Li, N. & Huang, F. Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry* 44, 4582-4590, doi:10.1021/bi047576b (2005).
40 Furuichi, Y. & Miura, K. A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253, 374-375 (1975).
41 Wei, C. M. & Moss, B. Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci U S A* 72, 318-322 (1975).
42 Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 5, 451-463, doi:10.1038/nrm1403nrm1403 [pii] (2004).

43 Celesnik, H., Deana, A. & Belasco, J. G. Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Mol Cell* 27, 79-90, doi:S1097-2765(07)00364-4 [pii]10.1016/j.molcel.2007.05.038 (2007).
44 Hopwood, D. A. *Genetic manipulation of Streptomyces: a laboratory manual*. (John Innes Foundation, 1985).
45 Neidhardt, F. C. I. J. L. S. M. *Physiology of the bacterial cell : a molecular approach*. (Sinauer Associates, 1990).

Chapter Three:

The Characterization of NAD-Linked RNA

Ye Grace Chen, Matt Edwards, David R Liu

Matt Edwards performed the alternate computational processing of Illumina Hi-Seq results.
Ye Grace Chen conducted and analyzed all of the other experiments described.

3.1 Introduction

We previously described the development of a nuclease-based screen that revealed NAD-linked RNA. Further characterization of NAD-linked RNA, including the size distribution and sequence, may elucidate its biological function. A more comprehensive understanding can add to the known functional diversity of RNA and contribute to a greater insight of the RNA world. In this chapter, we will describe the progress toward the characterization of NAD-linked RNA.

3.2 Size Distribution of NAD-Linked RNAs by Gel Electrophoresis

Size fractionation of total RNA followed by subsequent nuclease digestion and LC/MS analysis will yield the size distribution of small molecule-RNA conjugates. Determining the distribution of sizes of the NAD-linked RNA will reveal if there is promiscuous incorporation of NAD modification on the 5' end of total RNA or if there is localization of the modification to certain classes or size ranges of RNA. If NAD signal is detected at a basal level in all of the size fractionation samples, then there may be non-specific incorporation of NAD by *E. coli* polymerase *in vivo*. If NAD is found in only certain size ranges, then that may indicate the modification only exists on specific RNAs. Therefore, knowing the size distribution of NAD-linked RNA may yield insight into its biological function.

Prior to nuclease P1 digestion, size-exclusion chromatography removes molecules of molecular weight less than ~2,500 Da. To narrow the size window of NAD-linked RNAs and further fractionate the macromolecule fraction by size, we used silica-based RNA purification columns (Qiagen RNeasy columns). The columns separate RNA molecules into two fractions that are less than or greater than ~200 nucleotides in length (Fig. 3.1a). Each of the two fractions was then subjected to nuclease P1 digestion and LC/MS analysis.

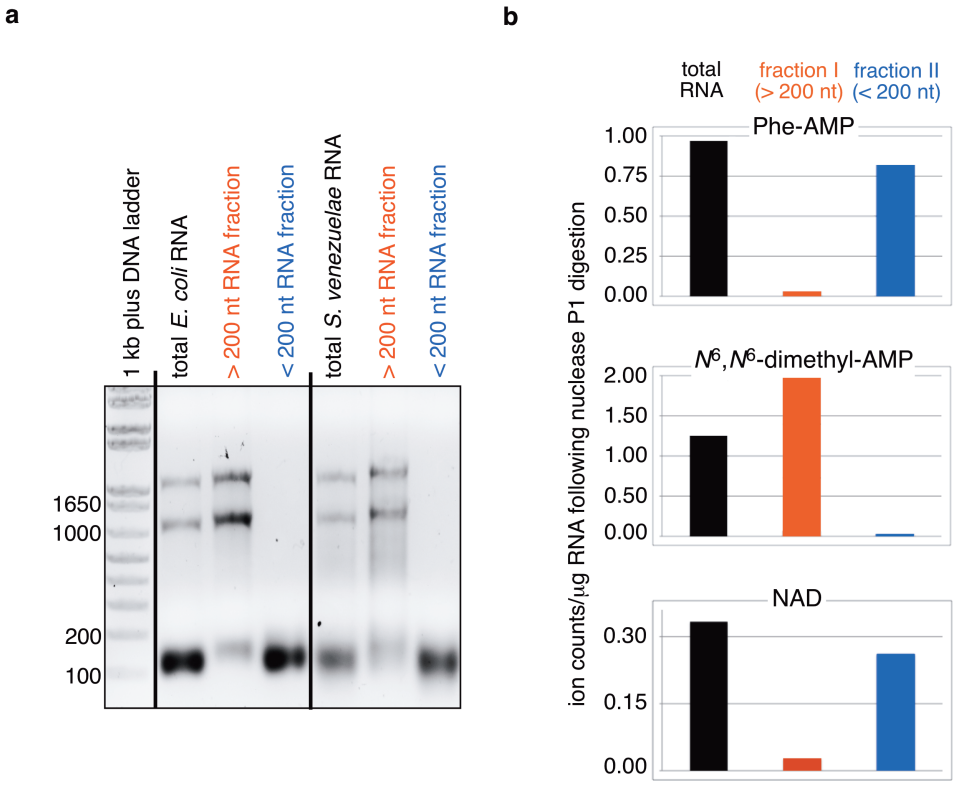


Figure 3.1. Silica column-based size fractionation of cellular RNA. (**a**) Total *E. coli* or *S. venezuelae* RNA was separated into fraction I (RNAs of length greater than ~200 nucleotides) and fraction II (RNAs of length less than ~200 nucleotides) using a Qiagen RNeasy silica column. The starting total RNA and each fraction (1 µg per lane) was analyzed by 1% TAE-agarose gel electrophoresis and stained with ethidium bromide. (**b**) Total *E. coli* RNA was separated into RNAs of length greater than 200 nt (fraction 1) and RNAs of length less than 200 nt (fraction II) using a silica column (Qiagen RNeasy). Each fraction was subjected to nuclease P1 digestion and analyzed by LC/MS. The presence of NAD in fraction II, similar to that of Phe-AMP, suggests that the NAD-linked RNA(s) are primarily less than 200 nt in length.

The rRNA nucleoside modification $N^6,N^{6'}$-dimethyladenine (conjugated to 1.5 kB-2.9 kB rRNAs) was detected in the greater than 200 base fraction and 3'-aminoacyl adenosine monophosphates conjugated to tRNAs (~76 nucleotides) were present predominantly in the less than 200 nucleotide flow-through fraction, as expected (Fig. 3.1b).[142,143] Like the tRNA modifications, the NAD-linked nucleotides were also predominantly detected in the flow-through RNA fraction (Fig. 3.1b). This result suggested that the NAD-linked RNA(s) from *E.*

coli and *S. venezuelae* are not widely distributed in their size but instead are below ~200 nucleotides in length, as was also observed in the case of CoA-linked RNA from the same organisms.[105] In addition, this finding further supported the hypothesis that the NAD modifications arise through a mechanism other than non-specific transcriptional initiation, which would be expected to generate a broad size distribution of NAD-linked RNAs.

3.3 Size Distribution of NAD-Linked RNAs by High Performance Liquid Chromatography

While the silica-based RNA purification columns were successful at separating total RNA into two fractions of greater than or less than 200 nucleotides in length, a finer resolution of NAD-linked RNA size would provide additional information. Therefore, to identify more precisely the length(s) of NAD-RNA, we developed a method to size fractionate total RNA by high performance liquid chromatography (HPLC) using an anion-exchange column. The method was first applied to a single-stranded RNA ladder of known lengths from 50 to 1000 nucleotides. The peaks could be separated by HPLC and gel electrophoresis detected bands of the appropriate sizes (Fig. 3.2a). Thus, separation of a known ladder validated the method to fractionate RNA of different sizes. We optimized the HPLC gradient for a wider range of oligonucleotide lengths and subjected total *E. coli* RNA to the protocol. Similar to the RNA ladder, an increase in RNA lengths in subsequent HPLC fractions was observed by gel electrophoresis. These results indicate that the HPLC method can size fractionate total RNA and that the quality of the RNA is preserved during the process (Fig. 3.2b).

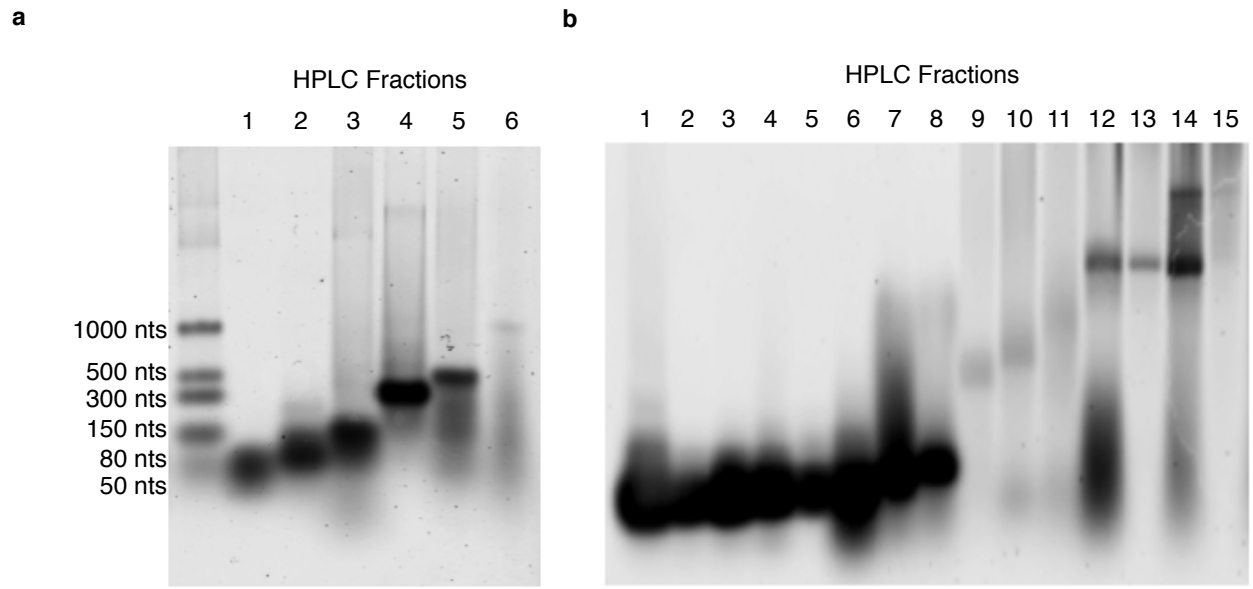


Figure 3.2. Size fractionation of RNA by high performance liquid chromatography (HPLC). (**a**) An RNA ladder can be separated by HPLC and (**b**) total *E. coli* RNA can be separated by HPLC. Each HPLC peak was collected, concentrated and analyzed by 1% TAE-agarose gel electrophoresis and stained with ethidium bromide.

To determine the size ranges of the NAD-linked RNAs, total RNA was subjected to the HPLC size fractionation protocol and five-minute fractions were collected during a 50-minute gradient. Since the loading capacity of the HPLC column is 50 μg of total RNA and we required at least 250 μg of total RNA in order to detect NAD signal from NAD-linked RNA, we collected and combined samples from at least five HPLC runs. The nucleic acids were concentrated and digested with nuclease P1 prior to LC/MS analysis.

The presence of rRNA and tRNA modifications in the corresponding HPLC fractions was a positive control for successful HPLC size fractionation. Indeed, rRNA modifications were detected in the 35 to 40 minute and 40 to 45 minute fractions and tRNA modifications and aminoacyl-adenylates in the five to ten minute fraction, as expected based on size information obtained from fractionation of the RNA ladder. The NAD daughter ion signal was predominantly detected in the five to ten minute fraction, the same one as tRNAs (Fig. 3.3).

Based on the retention times of the RNA ladder peaks, these results suggested that NAD-linked RNA is in the 30 to 120 nucleotides size range. This confirmed the previous silica-based column size fractionation result that NAD-RNA is less than 200 nucleotides long.

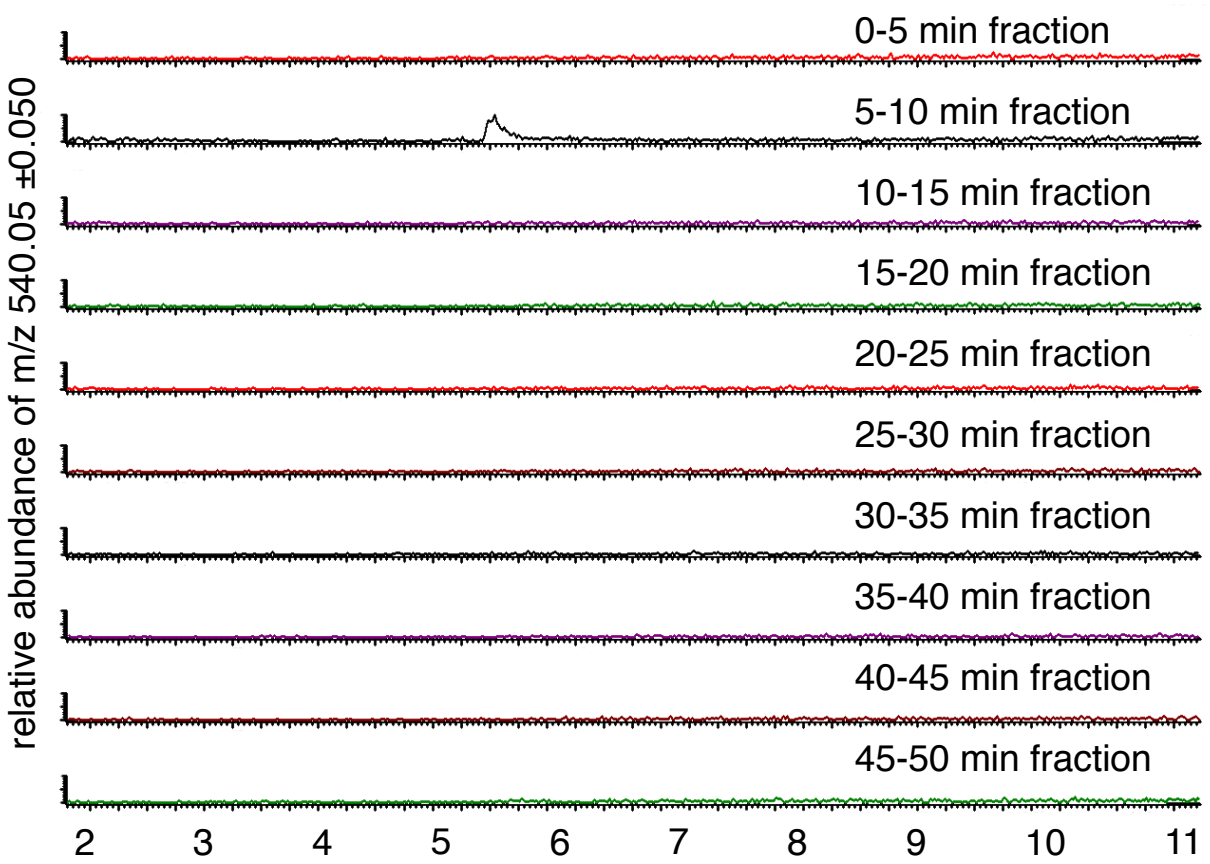


Figure 3.3. NAD-linked RNA is detected in the five to ten minute HPLC fraction, suggesting they are between 30 and 120 nucleotides in length based on fractionation of a RNA ladder. Following size fractionation by HPLC, the samples are concentrated, digested with nuclease P1 and analyzed by LC/MS.

One concern from the size fractionation data is that NAD-linked RNA may be distributed across the entire HPLC gradient at a low level so NAD signal may be undetectable in individual fractions. Since there is a decrease in overall signal before and after HPLC fractionation (due to material loss), the NAD signal in the five to ten minute sample may only be a portion of the complete NAD signal from total RNA. To determine the fraction of NAD signal found in the

five to ten minute fraction, the flow-through from the HPLC was collected, concentrated and digested prior to LC/MS analysis. The majority (~80%) of the NAD signal was detected in the five to ten minute fraction so most of the NAD-linked RNAs were in the 30 to 120 nucleotides size range. The NAD signal from samples that went through the HPLC size fractionation was also compared to a sample that did not go through the column to determine if there was significant loss of signal somewhere in the method and workup. ~70% of the NAD signal was detected by LC/MS analysis after size fractionation compared with no fractionation. This suggested that the RNA is not retained or stuck on the column after the HPLC size fractionation method is completed.

The HPLC gradient was adjusted further to allow for greater separation of the less than 200 nucleotide range from total RNA, focusing on the five to ten minute fraction where both NAD-linked RNA and tRNAs elute. Interestingly, the tRNA modifications and aminoacyl-adenylates continued to be detected in the same fraction as NAD-linked RNA. This suggested that NAD-linked RNAs may be a similar length as the tRNAs (which are typically ~76 nucleotides long).[142,143]

3.4 Enrichment of NAD-Linked RNA by Oxime Formation

An enrichment method to capture the small molecule-RNA conjugate for analysis is necessary in order to subject NAD-linked RNA to deep sequencing. To enrich for NAD-linked RNAs from total RNA, we envisioned enzymatically introducing a reactive chemical substituent selectively onto the nicotinamide of the RNA, and using bioorthogonal chemistry to add a biotin handle onto the chemically-modified NAD. The strong affinity of the biotin/streptavidin interaction will allow us to selectively pull out NAD-RNAs from the milieu of total RNA.[144]

After releasing the biotinylated NAD-RNA from the streptavidin, subsequent cloning and sequencing can be performed to determine the sequence(s) of NAD-RNA.

ADP ribosylcyclase converts NAD into cyclic adenine dinucleotide phosphate ribose (cADPR) *in vivo*; however, it has been demonstrated to replace the nicotinamide on NAD for other 3'-substituted pyridines *in vitro*.[145] To test the specificity of ADP ribosylcyclase, the four canonical nucleotides were subjected to the enzyme in the presence of 3'-pyridinecarboxyaldehyde. Each reaction contained the single nucleotide and the enzyme, or small molecule NAD and the enzyme. After the reaction, the small molecules were analyzed by LC/MS. Only the starting materials were detected in the reactions with the four canonical nucleotides and no products formed by an exchange reaction between the nucleotide bases and the 3'-pyridinecarboxyaldehyde were observed. When NAD is present, only the exchange product after the reaction was detected. To test that the enzyme is selective for NAD in a complex mixture, the nucleotides and NAD were pooled in one reaction. Similar to the individual reactions, only the expected exchange product where the 3'-pyridinecarboxyaldehyde was substituted in the place of nicotinamide on NAD was observed, and the canonical nucleotides remained unchanged. These experiments suggest that ADP ribosylcyclase is highly specific to exchanging the nicotinamide on NAD for 3'-substituted pyridines.

In addition to its selectivity, ADP ribosylcyclase was also an attractive enzyme for introduction of a reactive handle on NAD-linked RNA since the reaction requires acidic conditions. RNA is known to degrade under basic pH, where the rate is accelerated by stronger alkaline conditions that favor base catalysis.[146] Therefore, acidic conditions will better preserve the quality of RNA so more intact nucleic acids can be subjected to deep sequencing.

Given the high selectivity of ADP ribosylcyclase to exchange only the nicotinamide on small molecule NAD for 3'-pyridinecarboxyaldehyde, we envisioned introducing a reactive pyridine onto the NAD-linked RNA, which would then form a bond with a biotinylated reagent. Of the possible reactions that are compatible with biomolecules, oxime formation seems particularly attractive since the reaction has fast kinetics (with the addition of aniline)[147] and requires a low pH[148,149] so RNA quality should be preserved. Due to the more favorable reaction rates for the aldehyde compared with the ketone, we sought to use ADP ribosylcylase to exchange the nicotinamide with 3'-pyridinecarboxyaldehyde.[145,147] Since carboxyaldehyde is isosteric to some of the other substituents on pyridines that have been successfully exchanged by ADP ribosylcylcase, we anticipated that 3'-pyridinecarboxyaldehyde might be accepted as a substrate for exchange.[150] The crystal structure of the enzyme active site also shows room that may accommodate the RNA attachment to NAD. After the reaction of NAD-linked RNA with 3'-pyridinecarboxyaldehyde using ADP ribosylcyclase, introduction of biotin hydroxylamine to the aldehyde-exchanged NAD under oxime reaction conditions can incorporate biotin onto the NAD-RNA (Fig. 3.4).

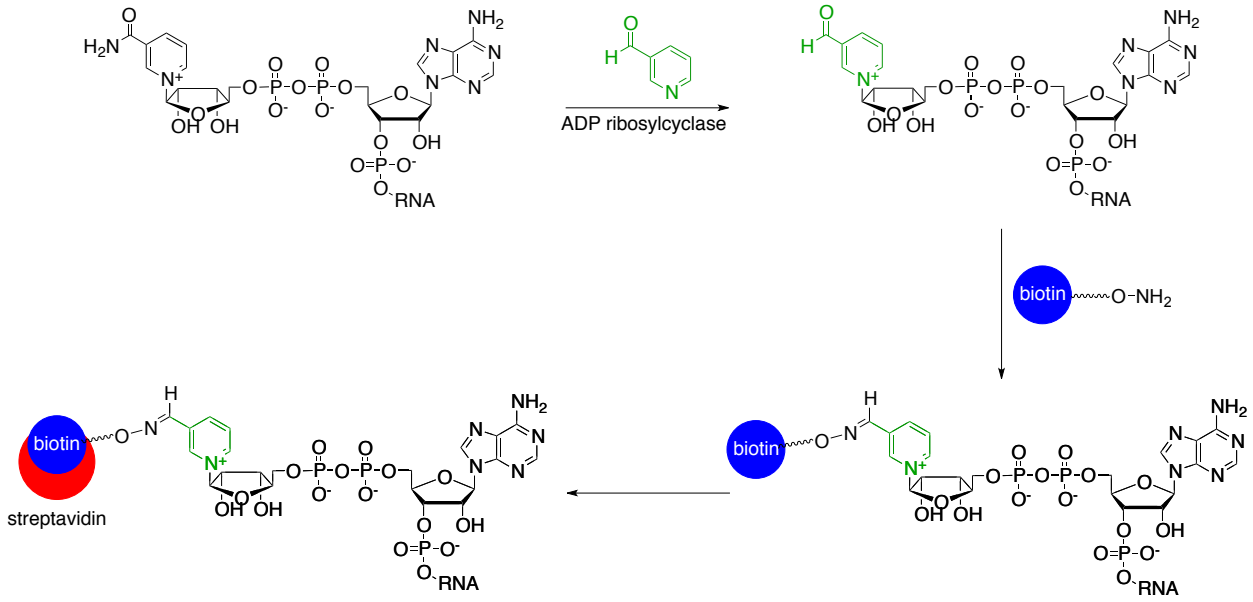


Figure 3.4. The method to introduce a biotin group onto NAD-linked RNA using oxime formation for enrichment of NAD-RNA. The nicotinamide is first exchanged for 3'-pyridinecarboxyaldehyde using ADP ribosylcyclase and then an oxime reaction is performed after introduction of a biotin hydroxylamine.

The exchange reaction was performed with 3'-pyridinecarboxyaldehyde on the small molecule NAD, authentic 5'-NAD-linked transcript and total RNA in quantitative yield. To monitor the reactions, the RNA was digested with nuclease and LC/MS analysis was conducted to scan for the desired nucleotide product (Fig. 3.5). Upon introduction of the biotin hydroxylamine, all three examples underwent the oxime reaction yielding a biotin linked to NAD. However, following the streptavidin pull-down, the desired NAD oxime product was not detectable in the elution. Most of the signal was found in the unbound fraction, suggesting that the streptavidin is not capturing the biotin oxime-NAD product. The incubation time for the RNA and streptavidin was increased, but the oxime product continued to be detected only in the flow-through.

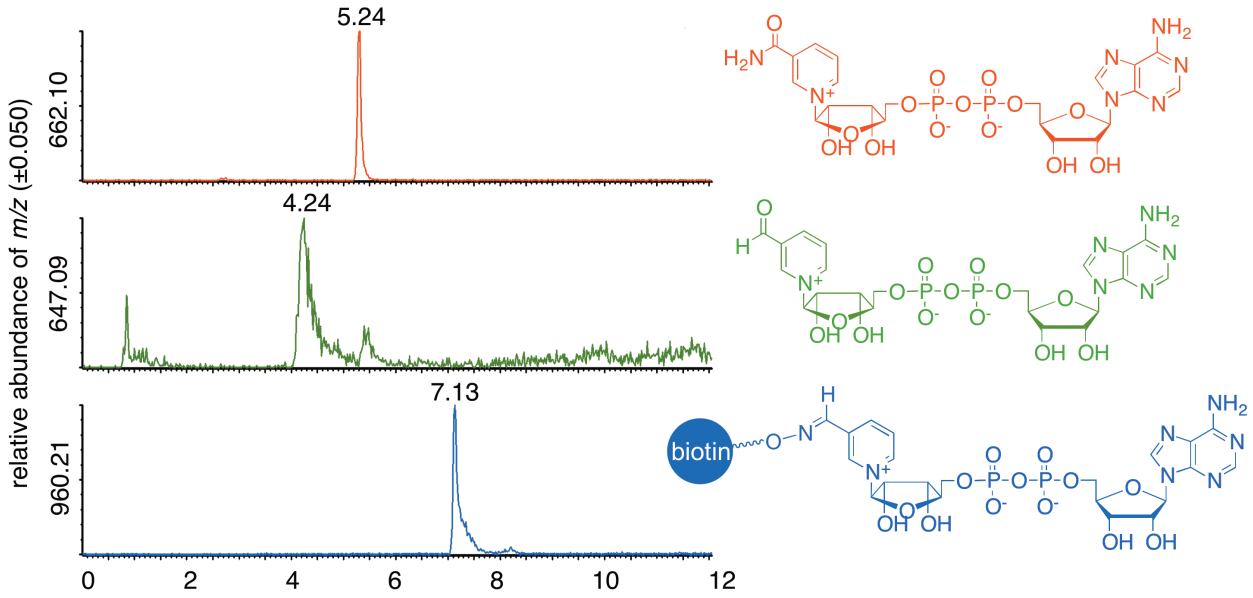


Figure 3.5. Small molecule NAD, *in vitro* transcribed NAD-linked RNA and cellular NAD-linked RNA can undergo the exchange reaction with 3'-pyridinecarboxyaldehyde and then an oxime reaction with biotin hydroxylamine.

We hypothesized that there may be off-target reactions occurring during the oxime reaction. If biotin from the hydroxylamine is incorporated nonspecifically on total RNA, the presence of all of the non-NAD-linked RNAs that are biotinylated may predominate the streptavidin binding. Thus, given the percentage and length of NAD-linked RNA, the NAD signal is expected to be detected primarily in the flow-through.

To test if hydroxylamine is reacting with more than just the aldehyde-exchanged NAD, we envisioned using a fluorescently-labeled hydroxylamine. We hypothesized that upon the introduction of an Alexa Fluor hydroxylamine, pyridinecarboxyaldehyde-exchanged NAD-linked RNA can form an oxime and gain a fluorescent tag (Fig. 3.6). The small molecule NAD, 5'-NAD-linked transcript and cellular RNA were subjected to the exchange and oxime reactions. The resulting nucleic acids were digested and analyzed by LC/MS, showing the expected exchange and oxime products in high yield. The *in vitro* transcribed NAD-linked RNA and

cellular RNA samples were also analyzed by polyacrylamide gel electrophoresis. As a negative control, the exchange reaction was omitted and the RNA was only subjected to the Alexa Fluor hydroxylamine under oxime reaction conditions. We expected to observe bands that correspond to the exchanged NAD-RNA and no signal in the unexchanged negative control lanes when imaging for Alexa Fluor-linked RNA. However, bands were detected in both sets of samples under the Alexa Fluor filter (Fig. 3.7), suggesting that there are off-target reactions occurring with the oxime formation.

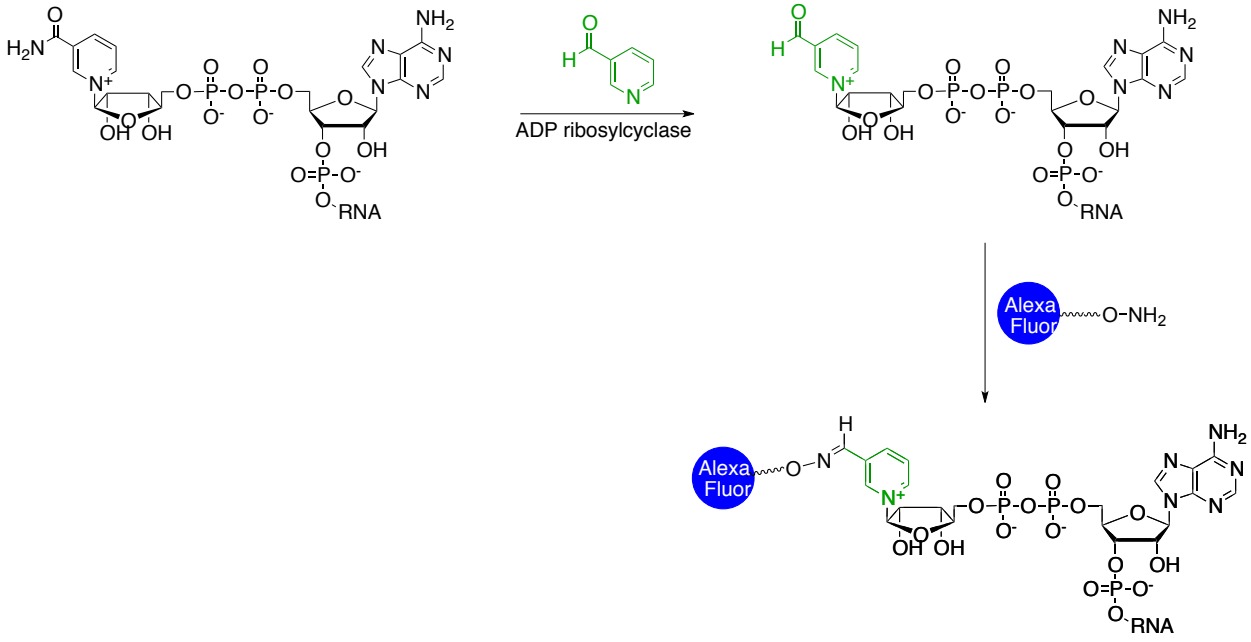


Figure 3.6. The method to add a fluorescent marker onto NAD-linked RNA for visualization of size distribution in total RNA. The nicotinamide on NAD is first exchanged for 3'-pyridinecarboxyaldehyde and then an oxime reaction is performed after introduction of an Alexa Fluor hydroxylamine.

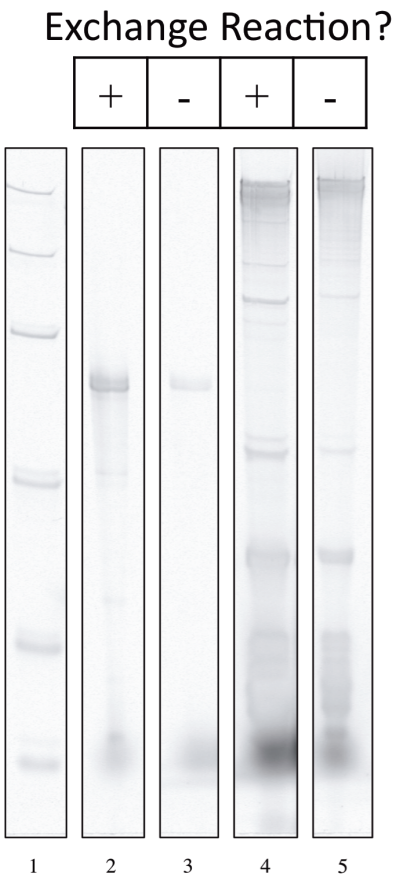


Figure 3.7. Oxime formation is not specific to reacting with the exchanged 3'-pyridinecarboxyaldehyde on *in vitro* transcribed NAD-linked RNA or cellular RNA. Lane 1: NEB low range ssRNA ladder, 2: *in vitro* transcribed 5'-NAD-linked RNA that was exchanged with 3'-pyridinecarboxyaldehyde, 3: *in vitro* transcribed 5'-NAD-linked RNA that was not exchanged with 3'-pyridinecarboxyaldehyde, 4: cellular RNA that was not exchanged with 3'-pyridinecarboxyaldehyde, 5: cellular RNA that was exchanged with 3'-pyridinecarboxyaldehyde.

To decrease unwanted reactions, the RNA was pre-blocked with hydroxylamine prior to the exchange reaction. However, the unexchanged RNA continued to produce signal on the gel imaged with the Alexa Fluor filter. The reaction time was decreased since we hypothesized that the aldehyde should be the most reactive group on RNA to the hydroxylamine and should form the oxime fastest. The yield was much lower for the exchanged NAD-linked RNA and the unexchanged still showed detectable signal under the Alexa Fluor filter.

To investigate the scope of off-target reactions that were occurring, total RNA that had reacted with 3'-pyridinecarboxyaldehyde and ADP ribosycyclase were subjected to biotin hydroxylamine under oxime formation conditions. As a negative control, total RNA that did not go through the exchange reaction was treated to the same oxime reaction conditions. Both sample were digested with nuclease and subjected to LC/MS analysis. Afterward, the data was processed by XCMS analysis to identify species that were enriched in the total RNA subjected to the exchange reaction compared with total RNA that had not gone through the exchange reaction. If the biotin hydroxylamine reacted specifically with the aldehyde on the exchanged NAD-linked RNA, there should only be one compound that is enriched. However, dozens of species appeared in both the unexchanged and exchanged reactions with hydroxylamine, so we started investigating other reactions that may react more specifically with the exchanged 5'-NAD-linked RNA.

3.5 Enrichment of NAD-Linked RNA by Click Chemistry

Of the other possible reactions that are compatible with biomolecules, the 1,3-dipolar cycloaddition or "click" reaction is especially favorable due to its compatibility with a wide range of functional groups, fast kinetics, and water solubility.[148,149] We envisioned exchanging the nicotinamide on NAD with an alkyne-substituted pyridine, which would be one of the click reaction partners.

Even though 3'-ethynylpyridine has not previously been shown as a substrate that ADP ribosylcyclase will accept to exchange the nicotinamide with, the crystal structure of the enzyme active site shows room for slight promiscuity so the exchange reaction might install an alkyne onto NAD.[150] A biotin azide can then be introduced to the EP-exchanged NAD-RNA, and in the

presence of copper, the two reagents should form a triazole, covalently linking the biotin to the NAD-RNA (Fig. 3.8).

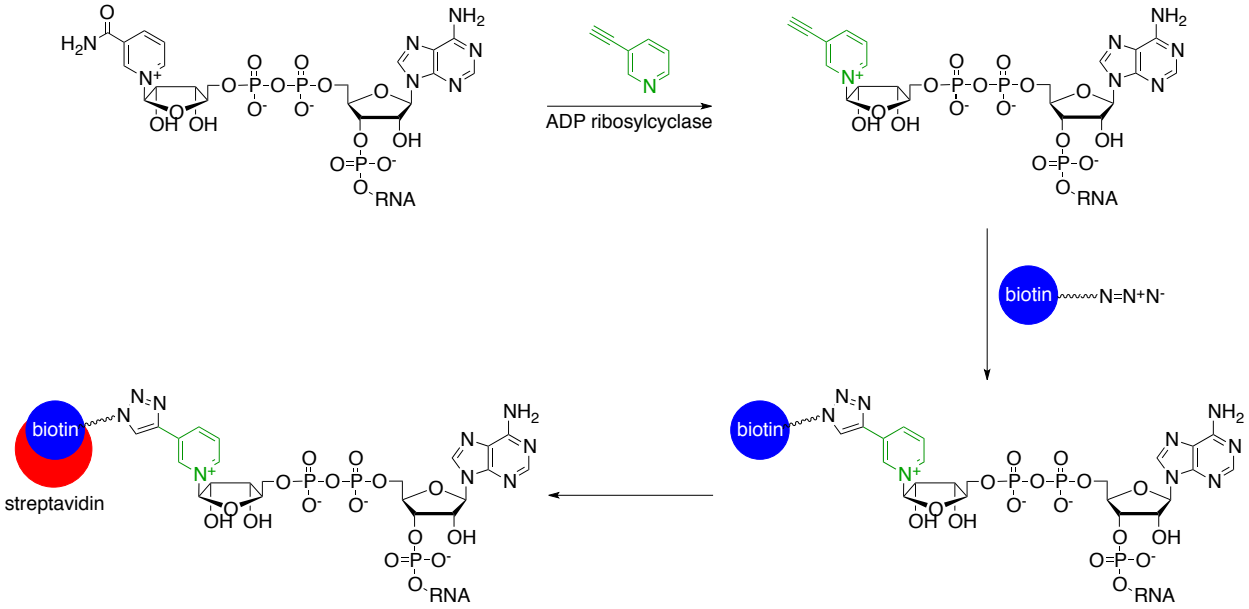


Figure 3.8. The method to add a biotin group onto NAD-linked RNA using click chemistry for enrichment of NAD-RNA. The nicotinamide is first exchanged for 3'-ethynylpyridine and then a click reaction is performed after introduction of a biotin azide.

The small molecule NAD, a 5'-NAD-linked transcript (generated *in vitro* using T7 RNA polymerase) and cellular RNA all performed the 3'-ethynylpyridine exchange reaction in quantitative yield. To monitor the reactions, the RNA was digested with nuclease, analyzed by LC/MS, and scanned for the presence of the desired nucleotide product. Upon exposure of a biotin azide, copper and ligand to the alkyne, the ethynylpyridine-exchanged small molecule and authentic NAD-linked RNA underwent click chemistry to produce a biotinylated ethynylpyridine-NAD. However, despite significant efforts to optimize the reaction for total RNA, the click products on total RNA was not reproducibly observable. Many conditions were screened, including $CuSO_4$ and ascorbic acid,[151,152] as well as Cu(I) without a reductant in the reaction (Figs. 3.9 and 3.10).[153,154] The 3'-ethynylpyridine-exchanged small molecule and NAD-

RNA showed the desired click products, but ethynylpyridine-exchanged total RNA only revealed the starting material. Furthermore, addition of *in vitro* generated ethynylpyridine-exchanged NAD-RNA into total RNA under click reaction conditions also did not yield product.

These results collectively indicate that there is something specific to the biological samples that inhibits the click reaction. We hypothesized that the total RNA may be sequestering the copper. Therefore, the concentration of the copper was increased, the concentration of RNA was decreased, and DMSO was added into the reaction to abrogate secondary structures. Unfortunately, none of these conditions produced the desired click product in the total RNA case, whereas the small molecule and authentic NAD-RNA transcript showed the biotinylated ethynylpyridine-NAD in all conditions.

NAD (μM, from RNA)	biotin (μM)	CuSO4 (μM)	THPTA (μM)	ascorbic acid (μM)
1	50	50	250	2500
0.5	50	50	250	2500
0.2	10	10	50	500
0.1	10	10	50	500
0.1	5	5	25	250
0.05	5	5	25	250
1	500	1000	2000	5000
0.2	100	200	400	1000
0.1	50	100	200	500
0.5	500	1000	2000	5000
0.1	100	200	400	1000
0.05	50	100	200	500
1	5000	10000	20000	5000
0.5	5000	10000	20000	5000
0.5	5000	10000	10000	50000

Figure 3.9. The click conditions attempted on small molecule NAD, *in vitro* transcribed NAD-linked RNA and cellular RNA with 10 mM PBS, pH 7.8 as the buffer.

NAD (μM, from RNA)	biotin (μM)	CuBr (μM)	THTA (μM)	DMSO (%)
0.5	500	1000	2000	5
0.5	500	1000	2000	10
0.1	100	200	400	10
0.05	50	100	200	10
0.2	100	200	400	10
0.1	50	100	200	10
0.5	500	1000	2000	50

Figure 3.10. The click conditions attempted on small molecule NAD, *in vitro* transcribed NAD-linked RNA and cellular RNA with 100 mM Tris-HCl, pH 8.0 as the buffer.

Through personal communication with Professor Subha Das from Carnegie Mellon who developed a protocol to perform click chemistry on the 5' end of RNA,[155] we learned that the traditional click ligands could inhibit the reaction in RNA. Acetonitrile was recommended as an alternative ligand to tris(hydroxypropyl)triazolylmethyl-amine (THPTA) and tris-(Benzyltriazolylmethyl)amine (TBTA). When acetonitrile was substituted for THPTA under click conditions, the exchanged small molecule NAD, authentic 5'-NAD-linked transcript and total RNA all produced the desired biotinylated NAD click product (Fig. 3.11). Yet, gel electrophoresis showed degradation of the authentic 5'-NAD-linked transcript and total RNA after the reaction. Degassing the solutions and decreasing reaction times did not significantly decrease the extent of RNA degradation as assessed by gel electrophoresis. However, since biotinylated NAD signal was still detectable after precipitation, we hypothesized that NAD-RNA remain a length that could be used for sequencing.

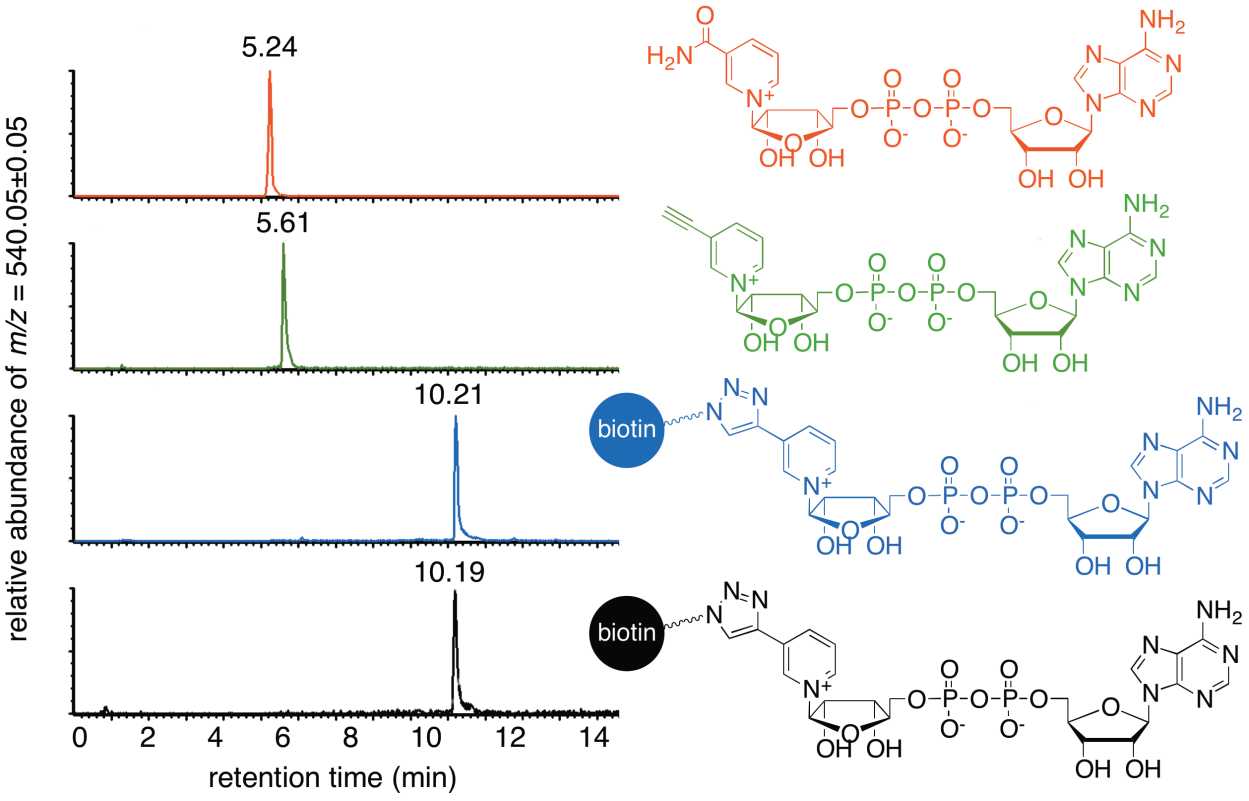


Figure 3.11. LC/MS spectra of *m/z* 540.05 after the exchange, click and streptavidin elution reactions showing successful product formation in high yield.

3.6 Enrichment of NAD-Linked RNA by Heck Reaction

The 1,3-dipolar cycloaddition or "click" reaction to specifically install a biotin group on the NAD-RNA results in degradation of total RNA, as observed by gel electrophoresis. While the NAD-linked RNA may still be intact enough to be accessible for adapter ligation and deep sequencing, we were interested in other reactions that could incorporate biotin onto NAD-RNA and preserve the RNA integrity. Professor Andrew Myers and coworkers had previously synthesized a biotin arylpalladium(II) trifluoroacetate complex that was used as a Heck coupling reagent to isolate the known protein targets of a series of natural products in affinity-isolation experiments (personal communication). This reagent allows biotin to be coupled to compounds containing a vinyl substituent under mild aqueous conditions.

To attach biotin to NAD-RNA conjugates, we envisioned first using ADP ribosylcyclase[145] to exchange the nicotinamide on NAD with 3'-vinylpyridine, which can be a Heck reaction partner, then exposing the resulting conjugate to the Myers biotin arylpalladium(II) reagent to attach biotin (Fig. 3.12). The vinylpyridine-exchange reaction was successfully performed on the small molecule NAD, an *in vitro* generated 5'-NAD-linked transcript, and NAD-RNA conjugates contained in isolated cellular RNA. In all three cases, LC/MS confirmed the presence of the expected modified nucleotide product. Upon exposure to the Myers biotin arylpalladium(II) reagent, the vinylpyridine-exchanged small molecule NAD, *in vitro*-transcribed NAD-RNA and total RNA all underwent Heck coupling chemistry to produce a biotinylated vinylpyridine-NAD.

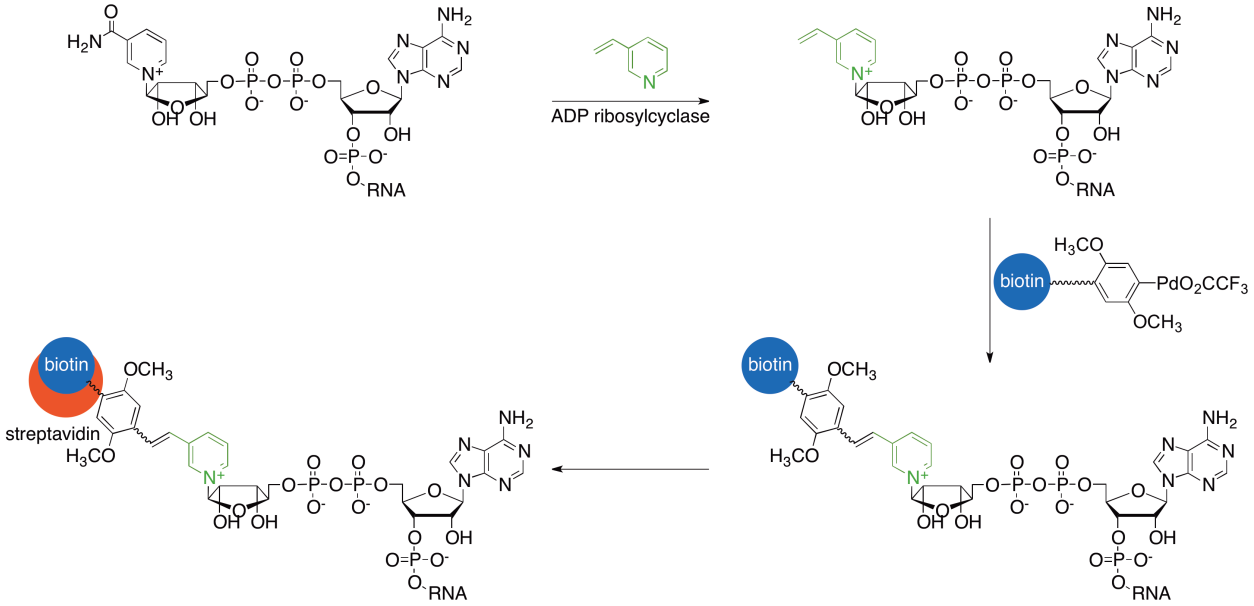


Figure 3.12. The method to add a biotin group onto NAD-linked RNA using the Heck reaction for enrichment of NAD-RNA. The nicotinamide is first exchanged for 3'-vinylpyridine and then a Heck reaction is performed after introduction of a biotin arylpalladium(II) trifluoroacetate complex.

The Myers lab reported that they achieved an average yield of ~50% conversion of vinylpyridine to biotinylated product in their cell lysate system (personal communication). The

reaction conditions were adjusted to optimize product yield, resulting in ~80 % of the expected product by LC/MS analysis. However, the reaction required almost an entire day, allowing for the potential of RNA degradation. Yet, under the two best reaction conditions (Fig. 3.13a), total RNA remained intact after both the vinylpyridine exchange and Heck coupling (Fig. 3.13b).

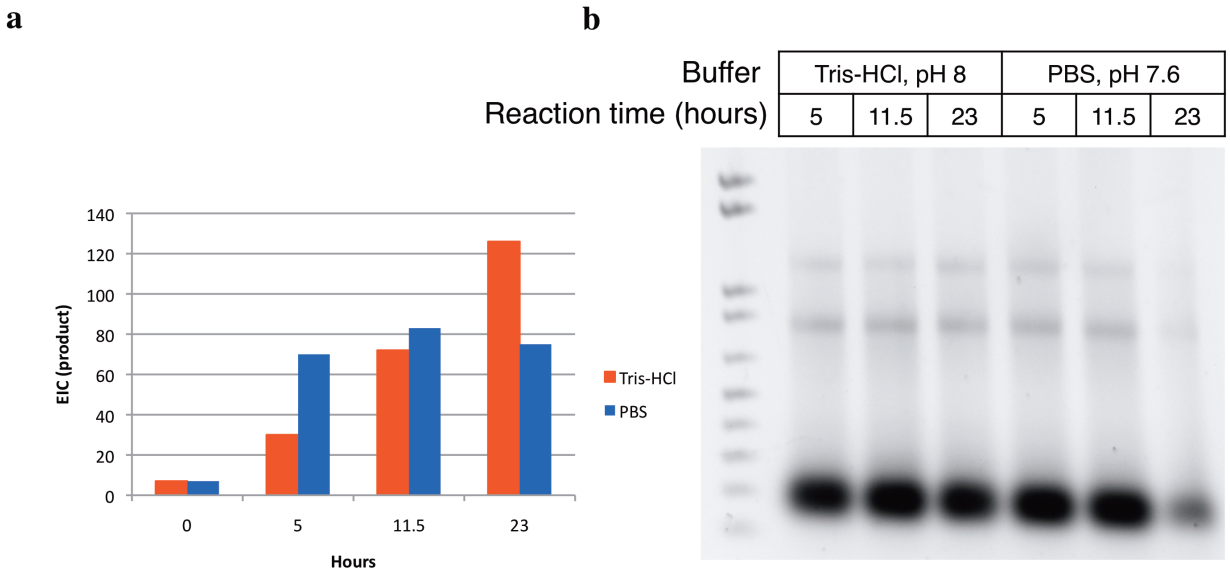


Figure 3.13. Total RNA can undergo the exchange reaction with 3'-vinylpyridine and the Heck coupling to product biotinylated product while maintaining integrity. (**a**) Total RNA was subjected to the exchange reaction with 3'-vinylpyridine and then the Heck reaction under two different buffer conditions. Nuclease digestion and LC/MS analysis followed, and the extracted ion chromatogram is reported. (**b**) Agarose gel electrophoresis of total RNA following the exchange reaction with 3'-vinylpyridine and the Heck reaction under two different buffer conditions showed that total RNA integrity is preserved during both reactions.

3.7 Development of Adapter Ligation Protocol for Illumina Deep Sequencing

The small RNA Illumina protocol guided the sequencing of the RNA(s) to which the small molecule is attached (Fig. 3.14). Total RNA was first treated with tobacco acid pyrophosphatase, which typically hydrolyzes the phosphoric acid anhydride bonds in the triphosphate bridge of the cap structure found in most eukaryotic mRNA, releasing the cap nucleoside and generating a 5′-monophosphorylated terminus on the RNA molecule.[156] The phosphatase also digests the triphosphate group at the 5′ end of prokaryotic transcripts,

generating an RNA molecule with a 5′-monophosphorylated terminus. We discovered that tobacco acid pyrophosphatase also acts on NAD-linked RNA that had undergone the exchange and click or Heck reactions to cleave the biotinylated substituent and revealing 5' monophosphates. Following preparation of the 5' end of RNA for ligation, the 3' adapter was added and then the 5' adapter using T4 RNA ligase I. After adapter ligation, reverse transcription and PCR amplified only the transcripts that contain both adapters.

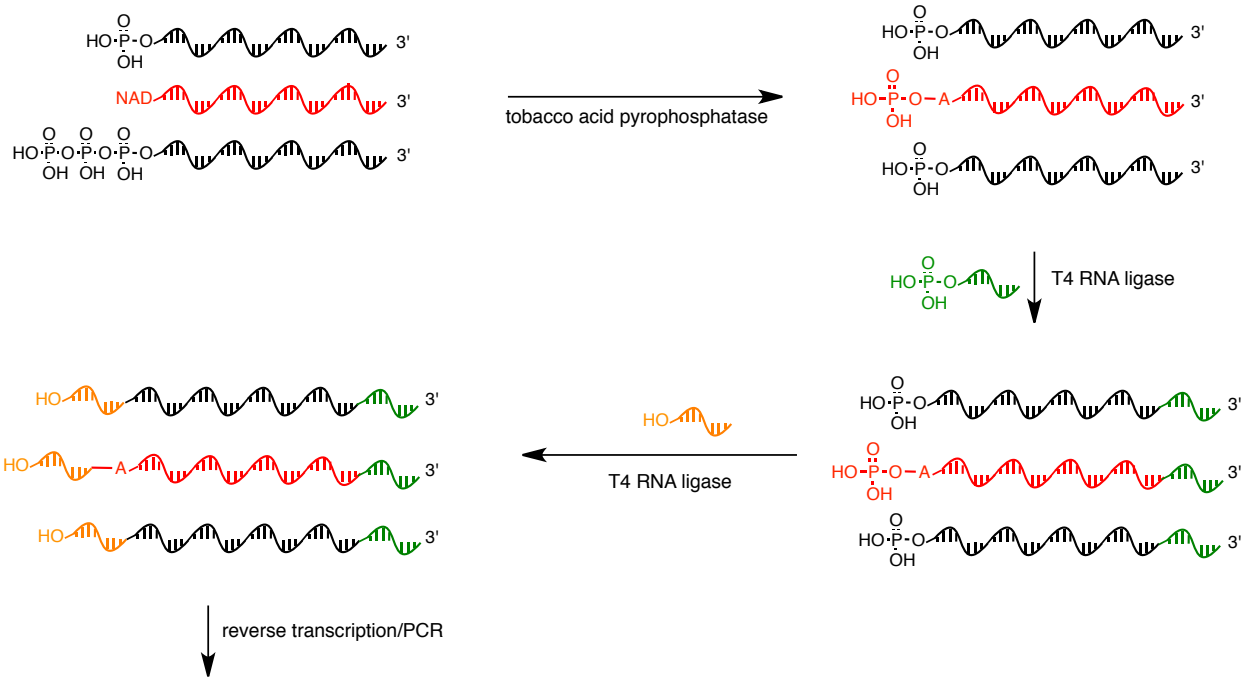


Figure 3.14. The general method to add adapters onto small RNA for Illumina deep sequencing, beginning with treatment of total RNA with tobacco acid pyrophosphatase followed by 3' and 5' adapter ligation. Reverse transcription and PCR will only amplify the RNAs containing both of the adapters.

To test the protocol, a 5'-NAD-linked transcript, generated *in vitro* using T7 RNA polymerase, was subjected to each of the steps. However, adapter ligation to the 228-mer template was not observed despite optimization of all of the reaction components. Instead, only the amplified 5' and 3' adapters ligation product was detected. When the positive control NAD-linked transcript was submitted to IDT's UNAFold program, the template was found to have a

high degree of secondary structure overall (Fig. 3.15). Specifically, the region near the 5' and 3' ends contain multiple stem loops, which might prevent adapter ligation from proceeding.

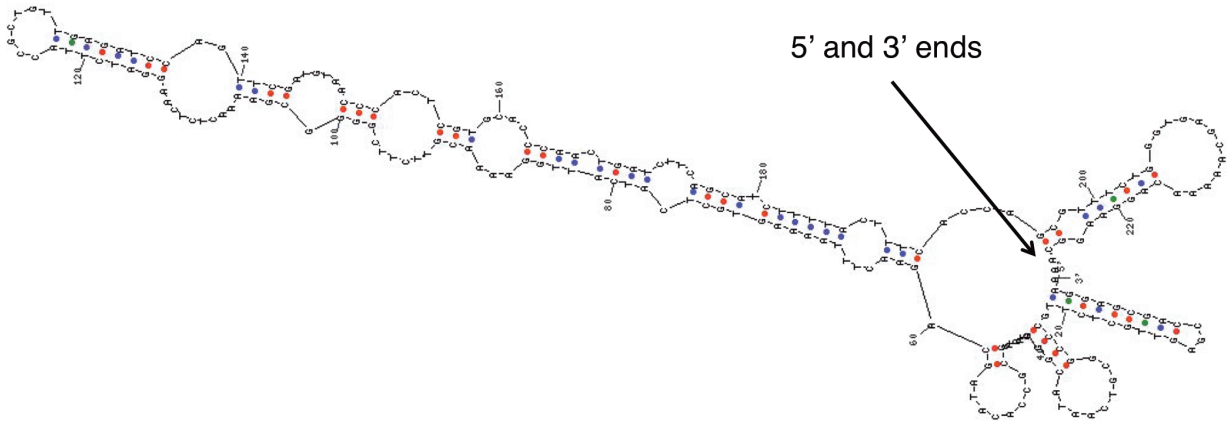


Figure 3.15. The secondary structure of *in vitro* generated 5'-NAD-linked RNA based on IDT UNAFold modeling suggesting that the 5' and 3' ends may be prevented from adapter ligation.

To generate a better positive control, the template was shortened to a 50-mer (Fig. 3.16) and submitted to the Illumina adapter ligation protocol (Fig. 3.14) This resulted in successful adapter ligation to the 50-mer template. However, the predominant product detected was the result from the amplification of the 5' and 3' adapters ligated to each together. To reduce the amount of ligated adapters present prior to sequencing, the cloning protocol was adjusted. Several iterations and optimization of the procedure resulted in the following protocol (Fig. 3.17). After the 3' adapter ligation, the reverse transcription primer was introduced to allow for hybridization to both the ligated and excess 3' adapter. Then, the 5' adapter ligation reaction was performed. Since T4 RNA ligase I only ligates single stranded nucleic acids, excess 3' adapter will not be accessible for ligation to the 5' adapter if the reverse transcription primer is hybridized. This prevented most of the unwanted 5' and 3' adapter ligations. Furthermore, gel purification following adapter ligation and PCR amplification ensured that the desired ligation

containing total RNA is the predominant product. The 50-mer RNA was subjected to each of these steps and the desired product without contamination was detected after PCR.

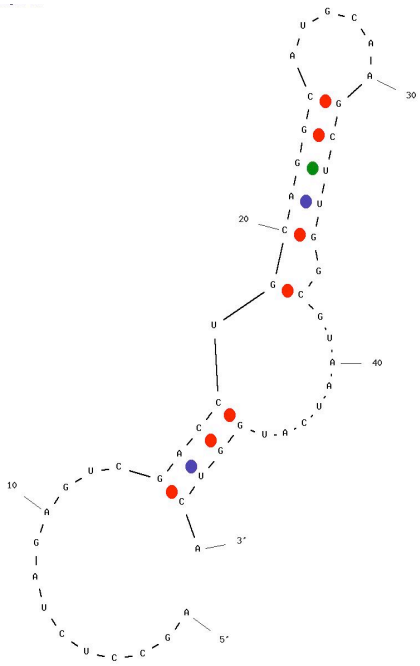


Figure 3.16. The secondary structure of a 50-mer RNA using IDT UNAFold modeling program suggesting the RNA may engage in adapter ligation.

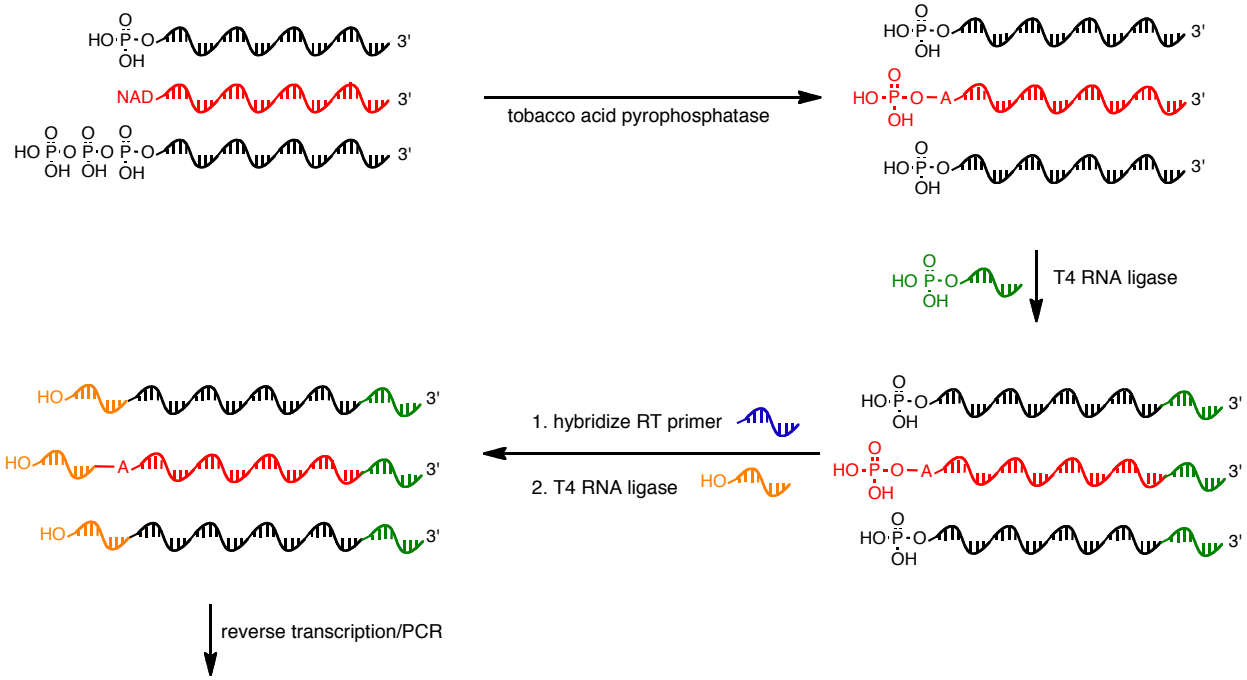


Figure 3.17. The modified cloning scheme to avoid adapter-adapter dimer formation for Illumina deep sequencing. After tobacco acid pyrophosphatase treatment and 3' adapter ligation, the reverse transcription primer is hybridized to the ligated and excess 3' adapter. 5' adapter ligation will only ligate onto single stranded nucleic acids, avoiding the undesired adapter-adapter dimer ligation. Reverse transcription and PCR should amplify just the desired ligation products.

3.8 Application of Adapter Ligation Protocol to Cellular RNA

E. coli total RNA samples was taken through the enrichment and modified adapter ligation protocols (Fig. 3.17) in preparation for Illumina high-throughput sequencing. The ADP ribosylcyclase-catalyzed exchange reaction was performed with either 3'-ethynylpyridine or 3'-vinylpyridine, and then biotin azide or the biotin arylpalladium(II) trifluoroacetate complex was introduced to attach biotin onto NAD-RNA (Fig. 3.18). Samples that did not go through the exchange reaction or the click or Heck reactions were prepared as negative controls. A portion of those samples was enriched by streptavidin pull-down. All of the comparative samples were treated with tobacco acid pyrophosphatase and taken through the adapter ligation protocol. A

barcode was included on the 3'-end of the 5' adapter for multiplexing purposes. Two biological replicates was subjected to each of these steps and submitted for Illumina deep sequencing.

Name	Exchange?	Biotinylation?	Streptavidin?	Barcode
Click_Pre	+	+	-	Mult1
Click_Post	+	+	+	Mult2
Heck_Pre	+	+	-	Mult1
Heck_Post	+	+	+	Mult2
EC click	-	+	-	Mult1
EP elute	+	-	+	Mult2
E	+	-	-	N6
EC heck	-	+	-	Mult1
VP elute	+	-	+	Mult2
V	+	-	-	N6

Figure 3.18. *E. coli* total RNA was subjected to different combinations of the exchange, biotinylation or streptavidin reactions to prepare comparative samples for Illumina deep sequencing.

3.9 Computation Analysis of Illumina Deep Sequencing Data Using the Tuxedo Platform

About 200 million reads was obtained from each lane containing comparative total RNA samples (Fig. 3.19). The sequences were subjected to the Tuxedo protocol[157] for alignment to the genome and identification of transcripts that are enriched following streptavidin pull-down. Prior to removing the barcodes, a very small percentage of the sequences aligned to the *E. coli* genome, as expected (Fig. 3.19).

Name	Number of Reads (Unfiltered)	% mapped to genome (unfiltered)	Number of Reads (filtered)	% mapped to genome (filtered)
Click1_Pre	101,306,061	0.37616999	38,240,355	37.74735156
Click1_Post	114,709,178	0.103256777	24,397,901	21.26935388
Click2_Pre	79,149,392	0.311062402	n/a	n/a
Click2_Post	103,375,261	0.131809099	n/a	n/a
Heck1_Pre	110,045,635	13.71993446	68,733,454	62.45904619
Heck1_Post	109,692,614	16.5774434	109,472,932	99.79972945
Heck2_Pre	111,876,426	12.32193367	81,261,175	72.63476132
Heck2_Post	154,514,760	19.77578551	103,286,633	66.8458036
EC1 click	101,918,771	5.530599461	n/a	n/a
EP1 elute	85,588,136	0.600056298	n/a	n/a
E1	46,927,860	5.112941438	n/a	n/a
EC2 click	107,863,347	3.567114416	25,058,904	23.23208457
EP2 elute	90,809,759	0.836596208	71,628,793	78.87785827
E2	35,095,837	6.392863632	19,466,219	55.46589187
EC1 heck	118,180,825	24.20982078	92,829,890	78.54902858
VP1 elute	100,515,355	10.12773111	98,573,243	98.06784546
V1	38,155,896	7.821145125	22,830,824	59.83563851

Figure 3.19. Data from Illumina deep sequencing of total RNA comparative samples using the Tuxedo platform.

After barcode removal, the sequences were mapped to the *E. coli* genome using TopHat, an alignment program for RNA-Seq reads. TopHat is built on the mapping program Bowtie, which aligns large sets of short DNA sequences to genomes, and is often used for processing sequences from Illumina Hi-Seq experiments.[158] While Bowtie cannot align reads to the genome that contain large gaps, TopHat solves this problem by breaking up reads that Bowtie cannot align on its own into smaller pieces, which can typically be aligned to the genome. Reads that align outside annotated genes are often strong evidence of new noncoding RNAs and protein-coding genes.[157] Alignments can also be used to accurately quantify gene and transcript expression, because the number of reads produced by a transcript is proportional to its abundance. Thus, the read densities can be used to measure transcript expression.

We mapped the biological replicates independently using TopHat, for more accurate statistical analysis. On average, about 62% of the sequences from samples that went through the exchange and corresponding click or Heck reactions aligned to the genome after barcode

removal. Sequences had to be longer than twelve bases with at most two mismatches in order to be counted as aligning to the genome. Unfortunately, TopHat was not able to process two of the Illumina sequencing lanes, which contained the second biological replicate of the click reaction and a replicate of the negative control with the click reaction.

Following sequence alignment to the genome, the abundance of transcripts between pre and post streptavidin was compared using the rest of the Tuxedo platform. The mapped reads from TopHat are provided as input to Cufflinks, a tool for transcriptome assembly and isoform quantitation from RNA-seq experiments. Cufflinks is a program that assembles aligned RNA-Seq reads into transcripts, estimates their abundances, and tests for differential expression and regulation transcriptome-wide for each replicate.[159] These assemblies are then merged with the reference transcriptome annotation into a unified annotation using Cuffmerge for further analysis. This merged assembly provides a uniform basis for calculating gene and transcript expression in each condition. The reads and the merged assembly are introduced to Cuffdiff, which calculates expression levels in two or more samples and tests the statistical significance of each observed change in expression between them.[157]

The Tuxedo platform should be able to identify transcripts that are enriched in the post-streptavidin pull-down samples compared with the pre-streptavidin samples, identifying possible candidates of RNA sequences that contain a 5' NAD modification. This analysis yielded 71 transcripts that are more than ten-fold enriched comparing before and after streptavidin pull-down. All of the enriched sequences are greater than 200 nucleotides in length and are from coding RNA. If one or all of the transcripts contain NAD on the 5' end, there is likely further processing of the transcript to end with a final product of NAD-linked RNA that is shorter than 200 nucleotides, as determined by our previous experiments.

Taking into account the enrichment and abundance following the streptavidin pull-down, the top three candidates for containing a 5' NAD on the transcript are the cca, adk and thrA transcripts (Fig. 3.20). The cca protein catalyzes the addition and repair of the essential 3'-terminal CCA sequence in tRNAs. Adenylate kinase (adk) is an essential enzyme that recycles AMP in active cells, converting ATP and AMP to two molecules of ADP and vice versa. The thrA protein has not been extensively characterized, but based on homology, it is predicted to have phosphatase activity.

Genome position	Pre-streptavidin counts	Post-streptavidin counts	Fold change	Protein Function (putative)
3313675-3314914	40.85	8018.06	196.26	catalyzes the addition and repair of the essential 3'-terminal CCA sequence in tRNAs
564415-565060	1132.5	41721.3	36.84	essential enzyme that recycles AMP in active cells; converts ATP and AMP to two molecules of ADP
4277857-4280743	216.62	7362.96	33.99	phosphatase

Figure 3.20. The top three enriched transcripts that should contain 5'-NAD-linked RNA based on the Tuxedo platform processing.

Antisense pull-downs were performed to verify if the identified RNA sequences contain NAD on the 5' end of the transcript. Antisense oligonucleotides were synthesized to different portions of the gene and cellular RNA was incubated with the probes. Given that the Tuxedo platform identifies entire transcripts that are enriched instead of the portions of transcripts that are enriched, probes were designed to cover entirely the identified transcripts for pull-down. The probes were spaced about two hundred nucleotides apart, and total RNA was incubated with each individual probe, as well as all of the probes for one transcript. Following release of the cellular RNA, nuclease P1 digestion and LC/MS analysis, NAD signal was not detected on the

pull-down transcripts. To confirm that the desired transcript was enriched, reverse-transcription/PCR was performed on the eluted material. Furthermore, to make sure that the NAD signal from a transcript is not being lost through the sample preparation, a probe to an *in vitro* generated 5'-NAD-linked transcript was designed, synthesized, and subjected to a pull-down. Following nuclease digestion and LC/MS analysis, NAD signal was detected only in the elution sample, and not in the unbound fraction. Taken together these results suggest that the transcripts we had identified using the Tuxedo platform do not contain NAD on their 5' end.

Comparison of the intensity of the signal from an individual NAD-linked RNA sequence to the total NAD signal from cellular RNA coupled with the pull-down efficiency will yield information about whether other sequences exist with the NAD modification. Furthermore, the pull-down will reveal the percentage of a sequence that contains the NAD modification, which will be useful in better understanding the potential cellular functions of NAD-linked RNA.

3.10 Reasons Why Cufflinks Did Not Accurately Identify NAD-Linked RNAs

While Cufflinks is capable of processing single-end RNA-Seq reads, the algorithms for transcript assembly and expression quantitation are much more accurate with paired-end reads.[157] The assembly algorithm explicitly handles paired end reads by treating the alignment for a given pair as a single object in the covering relation. This could distort the calculations, resulting in false positive identification of transcripts that are enriched in the post-streptavidin samples compared with the pre-streptavidin samples.

Sequencing read length is also a major consideration, and longer reads are generally preferable to short ones. Since NAD-linked RNA are in the 30 to 120 nucleotide length, a 50 bp read length should have been adequate. However, TopHat is more accurate when discovering

splice junctions with longer reads, and reads of 75 bp and longer are substantially more powerful than shorter reads.[157] The read length of our samples may have affected the higher than normal false positive rate in the computational processing.

3.11 Computation Analysis of Illumina Deep Sequencing Data Using Alternate Method

While the Tuxedo platform is relatively user friendly and readily available, the algorithms only calculate differential expression of entire transcripts. TopHat aligns the Illumina sequences to the genome and then Cufflinks/Cuffdiff identifies transcripts that are enriched from one sample to another. The algorithms do not identify the distribution of sequences within a transcript. However, since NAD-linked RNA may be a portion or a processed form of a transcript, a finer resolution than the entire transcript would be optimal.

To process the Illumina deep sequencing data independently from the previous method, the barcodes were removed using the FastX program. The parameters allowed for up to four mismatches in the barcode sequence and up to four shifts in order to achieve a partial match. The sequence length needed to be greater than ten bases for genome alignment. The short read aligner bwa was used to align the filtered reads to the *E. coli* reference genome K-12 MG1655 strain. As expected, the percentage of sequences that aligned to the genome after removing the barcode greatly increased (Fig. 3.21).

Name	Number of Reads (Unfiltered)	% mapped to genome (unfiltered)	Number of Reads (filtered)	% mapped to genome (filtered)
Click1_Pre	101,306,061	0.37616999	42,475,587	46.08437747
Click1_Post	114,709,178	0.103256777	58,840,050	44.6392607
Click2_Pre	79,149,392	0.311062402	40,971,529	50.9648814
Click2_Post	103,375,261	0.131809099	63,335,360	44.63335595
Heck1_Pre	110,045,635	13.71993446	58,383,099	36.1482641
Heck1_Post	109,692,614	16.5774434	85,139,089	30.85180869
Heck2_Pre	111,876,426	12.32193367	54,654,121	41.50623495
Heck2_Post	103,286,633	19.77578551	80,777,632	34.68182525
EC1 click	101,918,771	5.530599461	32,797,198	53.6578291
EP1 elute	85,588,136	0.600056298	47,218,985	48.96858981
E1	46,927,860	5.112941438	30,848,895	71.14651125
EC2 click	107,863,347	3.567114416	36,387,897	52.96607001
EP2 elute	90,809,759	0.836596208	47,251,012	54.94738906
E2	35,095,837	6.392863632	23,692,258	72.05784384
EC1 heck	92,829,890	24.20982078	66,659,883	52.90912577
VP1 elute	98,573,243	10.12773111	51,553,451	32.28392732
V1	38,155,896	7.821145125	25,703,747	69.16131992

Figure 3.21. Data from total RNA comparative samples using an alternate method to the Tuxedo platform.

Instead of using the Tuxedo platform for genome alignment and enrichment calculation based on transcripts, the sequences can first be aligned to each other, and then the enrichment can be calculated on a nucleotide resolution. The number of reads covering each base was counted, and then for each base, the counts of the 50 bases around it were summed. Finally, the fold-change difference was calculated for every tenth base. Since there were two biological replicates for each sample, the two replicates of the post-streptavidin samples were averaged and divided by the averages of the pre-streptavidin samples. For example, the average of the signal from the two click reactions that had gone through streptavidin pull-down were compared with the average of the signal from the two click reactions that had not gone through the pull-down. The regions where at least three consecutive points is about a fold threshold were counted as enriched since the size fractionation data suggested that NAD-linked RNA is greater than 30 nucleotides in length.

Using this method, 300 regions under the click reaction were enriched greater than 1.75-fold between the post and pre pull-down samples. 30 regions were enriched in the click reaction if the threshold was restricted to greater than 2-fold. For the Heck reaction, 952 regions were enriched greater than 1.75-fold and 372 regions were enriched greater than 2-fold. The biggest fold changes are 2.5 for the click samples and 3.4 for the Heck reaction.

When the enriched click and Heck transcript regions were compared with one another, two sequences were found within 100 bases from each other (Fig. 3.22). The two sequences are within transcripts that translate into a tRNA modification enzyme and a thiamin biosynthesis protein. The trmE modification enzyme has a 1,365 nucleotides transcript and is involved in the biosynthesis of the hypermodified nucleoside 5-methylaminomethyl-2-thiouridine, which is found in the wobble position of some tRNAs and affects ribosomal frameshifting. The enzyme is also involved in regulation of glutamate-dependent acid resistance. The second shared enriched region is from the sulfur carrier protein ThiS transcript, which is 201 nucleotides in length. This protein forms a complex with ThiF, ThiG, and ThiO to catalyze the formation of the thiazole moiety of thiamine pyrophosphate.

	Reaction	Begin	End	Enrichment of Post-Streptavidin/Pre-Streptavidin	Protein Function
1	Heck	3,983,059	3,983,239	2.501	Biosynthesis of the hypermodified nucleoside 5-methylaminomethyl-2-thiouridine, which is found in the wobble position of some tRNAs and affects ribosomal frameshifting
	Click	3,982,959	3,982,989	1.940	
2	Heck	4,290,247	4,290,327	2.715	With ThiF, ThiG, and ThiO catalyzes the formation of the thiazole moiety of thiamine pyrophosphate
	Click	4,290,252	4,290,292	1.858	

Figure 3.22. Shared regions in the click and Heck enriched samples that should contain 5'-NAD-linked RNA from the alternate method.

Antisense probes were designed for the two regions and pull-downs were performed on total RNA to confirm the presence of NAD on the identified transcript regions. After eluting from streptavidin and digesting with nuclease, LC/MS analysis showed no NAD signal in the elution fraction, while all of the expected NAD ion counts were present in the flow-through. This suggests that the two transcript regions predicted to contain the NAD modification did not have a detectable 5'-linked NAD. A positive control for the pull-down was an antisense probe to an *in vitro* transcribed NAD-linked RNA. LC/MS analysis following streptavidin elution and nuclease digestion showed NAD signal, suggesting that the pull-down and detection is working.

If the shared window is widened to 1000 bases, there are nineteen transcripts that are enriched in both click and Heck greater than 1.75 fold. The process of synthesizing biotinylated antisense probes to each of the shared enriched regions and performing the pull-downs is currently ongoing. Similarly, we are continuing the investigation of whether the transcript

regions that are uniquely enriched in either the click or Heck reaction contain the NAD modification.

3.12 Conclusions

Size fractionation of total RNA by silica-based columns that separate RNA greater than 200 nucleotides from the smaller fraction revealed NAD-linked RNA in the less than 200 nucleotides fraction. For a finer resolution, we developed a method to size fractionate total RNA using an anion-exchange column and high performance liquid chromatography (HPLC). Application of this method to *E. coli* cellular RNA generated a data set that contains other small molecule-RNA conjugates in addition to the NAD modification. Therefore, once researchers identify a novel small molecule-linked RNA, they can determine its sizes using the generated data allowing them to save time and reagents. LC/MS analysis of the HPLC size fractionation samples revealed that NAD-linked RNA is between 30 to 120 nucleotides in length. Both size fractionation methods resulted in a similar conclusion that NAD-linked RNA is of a similar size to tRNAs.

To enrich for NAD-linked RNA from a milieu of total RNA for deep sequencing, we developed methods that introduces a chemical handle onto NAD. ADP ribosycyclase is a specific enzyme that will selectively exchange the nicotinamide for a variety of 3'-substituted pyridines *in vitro*. The exchange reaction was performed to install a 3'-pyridinecarboxyaldehyde onto NAD-linked RNA. Then, introduction of a biotin hydroxylamine under oxime reaction conditions installed a biotin group onto the exchanged NAD-linked RNA. LC/MS analysis confirmed that the oxime reaction was successful at introducing a biotin group onto 3'-pyridinecarboxyaldehyde exchanged NAD-linked RNA. However, NAD-linked RNA was not

detectable after streptavidin elution. Gel electrophoresis of an oxime reaction between Alexa Fluor hydroxylamine and the 3'-pyridinecarboxyaldehyde exchanged NAD-linked RNA revealed that there were non-specific reactions between the hydroxylamine and RNA. Efforts to decrease or prevent the undesired reactions continued to result in NAD-linked RNA present in the flow-through after streptavidin binding.

The click reaction was a possible alternative to the oxime formation. We performed the exchange reaction with ADP ribosycyclase and 3'-ethynylpyridine, followed by introduction of biotin azide to install a biotin group onto NAD-linked RNA. Despite successfully biotinylating small molecule NAD and *in vitro* generated NAD-linked RNA, initial efforts to perform click chemistry on total RNA were ineffective. Screening of different conditions, including switching ligands, finally resulted in biotinylation of NAD-linked RNA in total RNA by the click reaction. However, while the desired biotinylated NAD was detectable after the click reaction and streptavidin binding and elution, gel electrophoresis revealed degradation of total RNA. Since copper is known to degrade RNA, other methods of bioorthogonal chemistry that were available for incorporating a biotin group onto NAD-linked RNA were investigated.

The Heck reaction was previously used to biotinylate a protein in cell lysate (personal communication). We subjected NAD-linked RNA to the exchange reaction with ADP ribosycyclase and 3'-vinylpyridine. Following the exchange, a biotin arylpalladium(II) trifluoroacetate complex that added a biotin group onto NAD-linked RNA was introduced under Heck reaction conditions. Reaction conditions were optimized to increase product yield, and despite the long reaction time of 23h, gel electrophoresis did not detect appreciable total RNA degradation.

To prepare comparative total RNA samples for Illumina deep sequencing, adapter ligation methods were developed. The main challenge to each of the attempted protocols was that the 5' and 3' adapters would ligate to each other and amplify more favorably than the samples containing total RNA ligated to the 5' and 3' adapters. To solve this problem, we employed hybridization of the reverse transcription primer to the excess 3' adapter prior to the 5' adapter ligation since the T4 RNA ligase only joins single stranded nucleic acids. This alteration coupled with gel purification was able to yield the desired total RNA ligated with the correct adapters as the predominant product. The developed method was applied to comparative total RNA samples subjected to click and Heck reaction conditions for Illumina deep sequencing.

Analysis of the Illumina deep sequencing results was first conducted using the Tuxedo platform, which consists of a genome alignment and enrichment calculation. The programs aligned the Illumina sequences to the *E. coli* genome and then counted the number of reads per transcript for all of the comparative samples. Enrichment between post-streptavidin and pre-streptavidin samples was calculated on a transcript level. Three candidates were highly enriched in both click and Heck reactions based on this analysis, so biotinylated antisense probes were designed and synthesized for pull-downs on *E. coli* total RNA. We did not observe NAD-linked RNA signal in the streptavidin elution fraction and only in the flow-through sample, suggesting that these three transcripts did not contain 5'-NAD-linked RNA.

When the pull-downs did not yield the expected NAD signal, another analysis independent of the Tuxedo platform was used. This time, the sequences were first aligned to each other and the coverage of each nucleotide was counted. The counts for every ten bases were summed, and the enrichment was calculated by dividing the average of the biological replicates of the post-streptavidin elution counts with the pre-streptavidin elution signal. These

calculations resulted in two regions that were enriched in both click and Heck samples. However, pull-downs using antisense probes to those regions did not result in detection of NAD-linked RNA in the elution fraction.

Efforts toward elucidating the sequence of NAD-linked RNA are currently ongoing. The sequence can generate additional hypothesis about the biological function of the small molecule-RNA conjugate, which may add to the functional diversity of RNA.

3.13 Materials and Methods

General. Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification.

RNA Size Fractionation by Silica Spin-Column. Total *E. coli* or *S. venezuelae* RNA was dissolved in 100 mM aqueous DTT and purified with Qiagen RNeasy kit (Qiagen) following the manufacturer's instructions. RNA from either the flow-through (less than ~200 nucleotides) or the eluted fraction (greater than ~200 nucleotides) was recovered by two consecutive precipitations with ethanol followed by desalting using NAP5 columns (GE Healthcare).

Size fractionation by HPLC. Size fractionation was performed using high performance liquid chromatography (HPLC, Agilent 1100) with an anion-exchange Dionex DNAPac PA200 4x250mm analytical column. Mobile phase A was 0.3M sodium perchlorate and mobile phase B was 20% acetonitrile in water. The flow rate was a constant 1.2mL/min, temperature was 80 °C and the mobile phase composition was as follows: linear increase from 30% A to 70% A for 5 min; linear increase over 35 min to 100% A; maintain at 100% A for 5 min before returning

linearly to 70% A over 2 min and then to 30% A over 3 min. Detection was carried out at 260 nm and 230 nm. 10ug of an RNA ladder (Ambion, Inc.) was injected to test the size fractionation method. 50ug of total RNA from *E. coli* was injected for each size fractionation run. Fractions were collected every 5 min and aggregated from four to ten runs. The first two fractions (0-5 min and 5-10 min) were concentrated using Millipore's Centriprep Centrifugal Filter Unit with Ultracel-3 membrane (YM-3) and the rest of the fractions were concentrated using the Centriprep Centrifugal Filter Unit with Ultracel-10 membrane (YM-10) by centrifugation (3,000 x g) until the total volume was less than 1.5mL. The retentate was lyophilized and ethanol precipitated prior to nuclease P1 digestion. For the digestion, the RNA was incubated with 2 U nuclease P1 (Wako Chemicals USA, Inc.) in 200uL of 50 mM NH4OAc, pH 4.5 at 37 ºC for 40 min. The digestion products were purified by size-exclusion chromatography (NAP5, GE Healthcare) and the small-molecule fraction was retained. The samples were resuspended in 18uL of milliQ water and injected for LC/MS analysis.

Cyclase Reaction. Nicotinamide was exchanged for 3'-substituted pyridine by incubating 2 mg of total RNA with 30 μL of 1M 3'-ethnylpyridine or 3'-vinylpyridine and 30uL of ADP ribosylcyclase (0.1-0.3U/uL) in 300 μL of 200 mM NH_4OAc, pH 4.5 at rt for 2h. The reaction was then extracted with acid-phenol chloroform (Ambion) twice, and the aqueous layer was washed once with chloroform. The aqueous layer was ethanol precipitated and the resulting pellet was dissolved in 50 mM NH_4OAc, pH 4.5, and subjected to size-exclusion chromatography using NAP5 columns (GE Healthcare).

Click Reaction. Small molecule NAD, *in vitro* generated 5'-NAD-linked transcript or total RNA containing NAD-linked RNA was subjected to click conditions as described in Table 3.1 and 3.2. The reagents were added in the following order: water, biotin azide (Invitrogen), DMSO (if included), ligand (Sigma-Aldrich), copper, and ascorbic acid (if included). After the reaction, the sample was ethanol precipitated, resuspended in water, and taken through size exclusion chromatography where the macromolecule fraction was collected.

Heck Reaction. Small molecule NAD, *in vitro* generated 5'-NAD-linked transcript or total RNA containing NAD-linked RNA was subjected to either Tris-HCl, pH 8 or PBS pH 7.8 as the buffer, with a final concentration of 0.1 mM biotinylated arylpalladium(II) complex (Fig. 3.23), 2% DMSO and 0.1% Triton. The reaction was sealed and incubated at rt for 1, 5, 11.5 and 23 h.

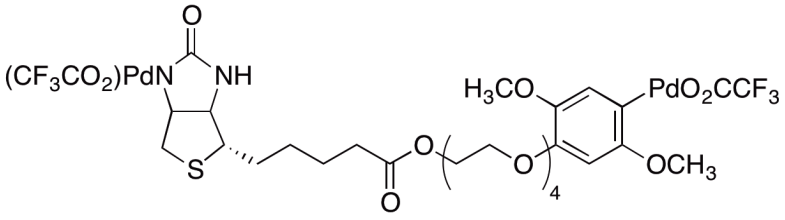


Figure 3.23. The biotinylated arylpalladium(II) complex used for the Heck coupling reaction.

Oxime Reaction. Nicotinamide was exchanged for 3'-pyridinecarboxyaldehyde by incubating 2 mg of total RNA with 2 µL of 3'-pyridinecarboxyaldehyde and 20 µL of ADP ribosylcyclase, 0.1-0.3 U/uL (Sigma-Aldrich) in 200 µL of 200 mM NH_4OAc, pH 4.5 at rt for 3h. The reaction was then extracted with acid-phenol chloroform (Ambion) twice, and the aqueous layer was washed once with chloroform. The aqueous layer was ethanol precipitated and the resulting pellet was dissolved in 95 µL of 100 mM anilinium acetate (made fresh prior to each reaction). 5 µL of 1 M biotin hydroxylamine in water was added to the reaction and incubated at rt with

shaking for 3h. The entire reaction was subjected to size-exclusion chromatography using NAP5 columns (GE Healthcare) and the macrocmolecular fraction was retained. For the polyacrylamide gel electrophoresis, Alexa Fluor hydroxylamine was substituted for biotin hydroxylamine. The reactions were ethanol precipitated and resuspended in water. A portion was loaded onto a 5% polyacrylamide gel and imaged using Typhoon Imager (GE Healthcare).

Adapter Ligation for Illumina Sequencing. 10 μg of total RNA was treated with 10 U of Tobacco Acid Pyrophosphatase (TAP, Epicenter) in 1x TAP buffer for 3 h at 37 °C. The sample was frozen and then lyophilized prior to resuspension in 1x T4 ssRNA ligase buffer, 1 μM 3' adapter, 1 mM ATP, and 10 U T4 ssRNA ligase (NEB). The reaction was incubated at 37 °C for 1 h. 1 μL of 50 mM reverse transcription primer was added to the reaction mixture and the sample was heated at 75 °C for five min, 37 °C for 30 min and then 25 °C for 15 min. With five minutes remaining in the previous incubation, heat the 5'-adapter (20 μM) at 70 °C for two minutes, then transfer to ice immediately. The ligation mixture contains 5 μM 5'-adapter, in addition to 1 mM ATP, and 10 U T4 ssRNA ligase (NEB). Incubate at 25 °C for 1 h. Take 3 μL of the ligation mixture and add to 1 mM dNTPs, 1x M-MuLV reverse transcriptase buffer, 100 mM DTT and 20 U of super RNase inhibitor (Ambion). Heat the sample at 42 °C for two minutes. Add 200 U of M-MuLV reverse transcriptase (NEB) and incubate at 42 °C for 1 h. For PCR amplification, take the reverse transcription reaction and add to 1 mM forward primer, 1 mM reverse primer, 1 mM dNTPs, 1x Phusion buffer and 2 U of Phusion DNA polymerase (Fermentas). The reaction was thermocycled at 98 °C for 30 sec, and then the following for a total of 20 cycles: 98 °C for 10 sec, 60 °C for 30 sec, 72 °C for 15 sec, and then 72 °C for ten minutes and holding at 12 °C.

Adapters:

Mult 1: GTTCAGAGTTCTACAGTCCGACGATCGATCGGAAGAGCACACGTNNNNNN
Mult 2: GTTCAGAGTTCUACAGTCCGACGATCACACTCTTTCCCTACACGNNNNNN

5' RNA Adapter
5' GUUCAGAGUUCUACAGUCCGACGAUC

3' RNA Adapter
5' P-UCGUAUGCCGUCUUCUGCUUGUidT

Samples in each lane for Illumina deep sequencing (Fig. 3.18)

1. Click1_Pre/Click1_Post mult1/mult2
2. Click2_Pre/Click2_Post mult1/mult2
3. Heck1_Pre/Heck2_Post mult1/mult2
4. Heck2_Pre/Heck2_Post mult1/mult2
5. E1/ EP1 elute/ EC1 click N6/mult2/mult1
6. E2/EP2 elute/EC2 click N6/mult2/mult1
7. V1/VP1 elute/EC1 heck N6/mult2/mult1

RT Primer
5' CAAGCAGAAGACGGCATACGA

Small RNA PCR Primer 1
5' CAAGCAGAAGACGGCATACGA

Small RNA PCR Primer 2
5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA

Small RNA Sequencing Primer
5' CGACAGGTTCAGAGTTCTACAGTCCGACGATC

50-mer:
AGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCA

228-mer
AGGAGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACAT
AGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGA
TCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAA
A

Computational Analysis of Illumina Sequencing

Barcode splitting. Read sequences were generated per lane on a standard Illumina flowcell. Reads from each lane were separated using the following molecular barcodes:

GATCGGAAGAGCACACGT
ACACTCTTTCCCTACAC

The FastX toolkit program fastx_barcode_splitter.pl was used with the given barcode sequences, with the parameters --mismatches 4 and --partial 4. These parameters allow for up to 4 mismatches in the barcode sequence and up to 4 shifts in order to achieve a partial match. These parameters were chosen in order to increase sensitivity for identifying error-prone barcode fragments.

Read filtering. All read sequences were examined with the quality control program FastQC. Frequent artifactual sequences were identified and removed. Specifically, the FastX toolkit program fastx_clipper was used to remove poly-A sequences and Illumina adapter sequences. The following sequences (or prefixes) were filtered:

TCGTATGCCGTCTTCTGCTTGTT	**Illumina Small RNA Adapter 2**
GATCGGAAGAGCACACGT	**Illumina Multiplexing Adapter 1**
ACACTCTTTCCCTACACG	**Illumina Multiplexing Adapter 2**

Reads were then trimmed to remove any barcode sequence and the unique 6-bp fingerprint. The trimming was accomplished with the FastX utility fastx_trimmer. Reads smaller than 10-bp after this filtering were excluded from further analysis.

Read alignment. Filtered reads were aligned to the *E. coli* reference genome, K-12 MG1655 strain (NCBI accession NC_000913.2) using the short read aligner bwa.[160] The default settings for short single-ended read alignment were used, with multi-threading mode enabled.

Pileup. Read alignments were converted into genome pileup files using the SAMtools utility suite.[161] These files list the sequenced bases observed at each position of the reference genome. The samtools depth command was used to calculate the number of reads covering each base pair position in the *E. coli* genome for each sequenced dataset.

Statistics. A custom script was used to calculate the read coverage of 50-bp sliding windows, calculated in 10-bp intervals along the genome. This computation was performed for each sequenced dataset. Read coverage fold changes were calculated for each replicated condition and compared to the appropriate replicated control experiment. For a given fold change threshold, contiguous regions (spaced 10-bp apart) of at least three significant segments were identified. Regions surpassing a given threshold (either 1.75x or 2x) were identified for each experimental condition.

FastX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html)

FastQC utility (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

3.14 References:

1 Palade, G. E. & Siekevitz, P. PANCREATIC MICROSOMES. *The Journal of Biophysical and Biochemical Cytology* **2**, 671-690, doi:10.1083/jcb.2.6.671 (1956).

2 Palade, G. E. & Porter, K. R. STUDIES ON THE ENDOPLASMIC RETICULUM. *The Journal of Experimental Medicine* **100**, 641-656, doi:10.1084/jem.100.6.641 (1954).

3 Porter, K. R. & Palade, G. E. STUDIES ON THE ENDOPLASMIC RETICULUM. *The Journal of Biophysical and Biochemical Cytology* **3**, 269-300, doi:10.1083/jcb.3.2.269 (1957).

4 Siekevitz, P. & Palade, G. E. A Cytochemical Study on the Pancreas of the Guinea Pig. *The Journal of Biophysical and Biochemical Cytology* **4**, 309-318, doi:10.1083/jcb.4.3.309 (1958).
5 Palade, G. E. & Siekevitz, P. LIVER MICROSOMES. *The Journal of Biophysical and Biochemical Cytology* **2**, 171-200, doi:10.1083/jcb.2.2.171 (1956).
6 Zamecnik, P. C. & Keller, E. B. RELATION BETWEEN PHOSPHATE ENERGY DONORS AND INCORPORATION OF LABELED AMINO ACIDS INTO PROTEINS. *Journal of Biological Chemistry* **209**, 337-354 (1954).
7 Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **231**, 241-257 (1958).
8 Crick, F. H. On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138-163 (1958).
9 Brenner, S., Jacob, F. & Meselson, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **190**, 576-581 (1961).
10 Bishop, J. M. *et al.* The low molecular weight RNAs of Rous sarcoma virus: II. The 7 S RNA. *Virology* **42**, 927-937 (1970).
11 Walker, T. A., Pace, N. R., Erikson, R. L., Erikson, E. & Behr, F. The 7S RNA Common to Oncornaviruses and Normal Cells is Associated with Polyribosomes. *Proceedings of the National Academy of Sciences* **71**, 3390-3394 (1974).
12 Zieve, G. & Penman, S. Small RNA species of the HeLa cell: Metabolism and subcellular localization. *Cell* **8**, 19-31 (1976).
13 Walter, P., Ibrahimi, I. & Blobel, G. n. Translocation of Proteins across the Endoplasmic Reticulum I. Signal Recognition Protein (SRP) Binds to In-vitro-Assembled Polysomes Synthesizing Secretory Protein. *The Journal of Cell Biology* **91**, 545-550 (1981).
14 Walter, P. & Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**, 691-698 (1982).
15 Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. A possible precursor of immunoglobulin light chains. *Nature: New biology* **239**, 117-120 (1972).
16 Keiler, K. C. Biology of trans-Translation. *Annual Review of Microbiology* **62**, 133-151, doi:10.1146/annurev.micro.62.081307.162948 (2008).
17 Ray, B. K. & Apirion, D. Characterization of 10S RNA: A new stable RNA molecule from <i>Escherichia coli</i&gt. *Molecular and General Genetics MGG* **174**, 25-32, doi:10.1007/bf00433301 (1979).
18 Tyagi, J. S. & Kinger, A. K. Identification of the 10Sa RNA structural gene of Mycobacterium tuberculosis. *Nucleic Acids Research* **20**, 138, doi:10.1093/nar/20.1.138 (1992).
19 Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proceedings of the National Academy of Sciences* **91**, 9223-9227 (1994).
20 Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. H. A. mut-7 of C. elegans, Required for Transposon Silencing and RNA Interference, Is a Homolog of Werner Syndrome Helicase and RNaseD. *Cell* **99**, 133-141 (1999).
21 Tabara, H. *et al.* The rde-1 Gene, RNA Interference, and Transposon Silencing in C. elegans. *Cell* **99**, 123-132 (1999).

22 Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11**, 1017-1027 (2001).
23 Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366, doi:http://www.nature.com/nature/journal/v409/n6818/suppinfo/409363a0_S1.html (2001).
24 Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. *Cell* **101**, 25-33 (2000).
25 Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development* **15**, 188-200, doi:10.1101/gad.862301 (2001).
26 Bass, B. L. Double-Stranded RNA as a Template for Gene Silencing. *Cell* **101**, 235-238 (2000).
27 Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215-233 (2009).
28 Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906 (2000).
29 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
30 Lee, R. C. & Ambros, V. An Extensive Class of Small RNAs in Caenorhabditis elegans. *Science* **294**, 862-864, doi:10.1126/science.1065329 (2001).
31 Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, 858-862, doi:10.1126/science.1065062294/5543/858 [pii] (2001).
32 Brodersen, P. *et al.* Widespread Translational Inhibition by Plant miRNAs and siRNAs. *Science* **320**, 1185-1190, doi:10.1126/science.1159151 (2008).
33 Seto, A. G., Kingston, R. E. & Lau, N. C. The Coming of Age for Piwi Proteins. *Molecular Cell* **26**, 603-609 (2007).
34 Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246-258 (2011).
35 Brantl, S. Antisense-RNA regulation and RNA interference. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1575**, 15-25 (2002).
36 Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **81**, null, doi:doi:10.1146/annurev-biochem-051410-092902 (2012).
37 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:321/5891/960 [pii]10.1126/science.1159689 (2008).
38 Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem Biol* **9**, 1043, doi:S1074552102002247 [pii] (2002).
39 Mironov, A. S. *et al.* Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**, 747-756 (2002).
40 Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952-956 (2002).

41 Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences* **99**, 15908-15913, doi:10.1073/pnas.212628899 (2002).
42 Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences* **29**, 11-17 (2004).
43 Nou, X. & Kadner, R. J. Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proceedings of the National Academy of Sciences* **97**, 7190-7195, doi:10.1073/pnas.130013897 (2000).
44 Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends in Genetics* **15**, 439-442 (1999).
45 Stormo, G. D. & Ji, Y. Do mRNAs act as direct sensors of small molecules to control their expression? *Proceedings of the National Academy of Sciences* **98**, 9465-9467, doi:10.1073/pnas.181334498 (2001).
46 Gold, L. *et al.* From oligonucleotide shapes to genomic SELEX: Novel biological,Äâregulatory,Äâloops. *Proceedings of the National Academy of Sciences* **94**, 59-64 (1997).
47 Woese, C. *The genetic code*. (Harper and Row, New York, 1967).
48 Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157 (1982).
49 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
50 Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228, doi:10.1038/418222a (2002).
51 Perrotta, A. T., Shih, I.-h. & Been, M. D. Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme. *Science* **286**, 123-126, doi:10.1126/science.286.5437.123 (1999).
52 Santoro, S. W., Joyce, G. F., Sakthivel, K., Gramatikova, S. & Barbas, C. F. RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. *Journal of the American Chemical Society* **122**, 2433-2439, doi:10.1021/ja993688s (2000).
53 Wyatt, G. R. Occurrence of 5-methylcytosine in nucleic acids. *Nature* **166**, 237-238 (1950).
54 Limbach, P. A., Crain, P. F. & McCloskey, J. A. Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22**, 2183-2196 (1994).
55 Gott, J. M. in *Methods in enzymology* (Academic Press/Elsevier, 2007).
56 Czerwoniec, A. *et al.* MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Research* **37**, D118-D121, doi:10.1093/nar/gkn710 (2009).
57 Motorin, Y. & Grosjean, H. in *eLS* (John Wiley & Sons, Ltd, 2001).
58 Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-351, doi:S0968000402021096 [pii] (2002).
59 Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **289**, 920-930, doi:10.1126/science.289.5481.920 (2000).
60 Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 √Ö Resolution. *Science* **289**, 905-920, doi:10.1126/science.289.5481.905 (2000).

61 Harms, J. *et al.* High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. *Cell* **107**, 679-688 (2001).
62 Yusupov, M. M. *et al.* Crystal Structure of the Ribosome at 5.5 √Ö Resolution. *Science* **292**, 883-896, doi:10.1126/science.1060089 (2001).
63 Kitchingman, G. R. & Fournier, M. J. Modification-deficient transfer ribonucleic acids from relaxed control Escherichia coli: structures of the major undermodified phenylalanine and leucine transfer RNAs produced during leucine starvation. *Biochemistry* **16**, 2213-2220, doi:10.1021/bi00629a027 (1977).
64 Grosjean, H. in *Topics in current genetics,; 12;* (New York, 2005).
65 Persson, B. C., Gustafsson, C., Berg, D. E. & Bj√∂rk, G. R. The gene for a tRNA modifying enzyme, m5U54-methyltransferase, is essential for viability in Escherichia coli. *Proceedings of the National Academy of Sciences* **89**, 3995-3998 (1992).
66 Anderson, J. *et al.* The essential Gcd10p‚ÄìGcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes & Development* **12**, 3650-3662, doi:10.1101/gad.12.23.3650 (1998).
67 Gerber, A. P. & Keller, W. An Adenosine Deaminase that Generates Inosine at the Wobble Position of tRNAs. *Science* **286**, 1146-1149, doi:10.1126/science.286.5442.1146 (1999).
68 Gu, W., Jackman, J. E., Lohan, A. J., Gray, M. W. & Phizicky, E. M. tRNAHis maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5‚Äô end of tRNAHis. *Genes & Development* **17**, 2889-2901, doi:10.1101/gad.1148603 (2003).
69 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *The Biochemical journal* **55**, 774-782 (1953).
70 Cohn, W. E. & Volkin, E. Nucleoside-5[prime]-Phosphates from Ribonucleic Acid. *Nature* **167**, 483-484 (1951).
71 Grosjean, H. Vol. 12 *Topics in Current Genetics* (ed Henri Grosjean) 1-22 (Springer Berlin / Heidelberg, 2005).
72 Buck, M., Connick, M. & Ames, B. N. Complete analysis of tRNA-modified nucleosides by high-performance liquid chromatography: The 29 modified nucleosides of Salmonella typhimurium and Escherichia coli tRNA. *Analytical Biochemistry* **129**, 1-13 (1983).
73 Klawitter, J., Schmitz, V., Klawitter, J., Leibfritz, D. & Christians, U. Development and validation of an assay for the quantification of 11 nucleotides using LC/LC‚Äìelectrospray ionization‚ÄìMS. *Analytical Biochemistry* **365**, 230-239 (2007).
74 Tuytten, R., Lemière, F., Dongen, W. V., Esmans, E. L. & Slegers, H. Short capillary ion-pair high-performance liquid chromatography coupled to electrospray (tandem) mass spectrometry for the simultaneous analysis of nucleoside mono-, di- and triphosphates. *Rapid Communications in Mass Spectrometry* **16**, 1205-1215, doi:10.1002/rcm.704 (2002).
75 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618-618 (1986).
76 Scott, A. I. How were porphyrins and lipids synthesized in the RNA world? *Tetrahedron Letters* **38**, 4961-4964 (1997).
77 Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci U S A* **86**, 7054-7058 (1989).

78 Gartner, Z. J., Kanan, M. W. & Liu, D. R. Multistep small-molecule synthesis programmed by DNA templates. *Journal of the American Chemical Society* **124**, 10304-10306 (2002).
79 Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. & Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **431**, 545-549, doi:10.1038/nature02920 (2004).
80 Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *Journal of the American Chemical Society* **123**, 6961-6963 (2001).
81 Zev J. Gartner, M. W. K. D. R. L. Expanding the Reaction Scope of DNA-Templated Synthesis. *Angewandte Chemie International Edition* **41**, 1796-1800 (2002).
82 Rozenman, M. M., Kanan, M. W. & Liu, D. R. Development and Initial Application of a Hybridization-Independent, DNA-Encoded Reaction Discovery System Compatible with Organic Solvents. *Journal of the American Chemical Society* **129**, 14933-14938 (2007).
83 Momiyama, N., Kanan, M. W. & Liu, D. R. Synthesis of acyclic alpha,beta-unsaturated ketones via Pd(II)-catalyzed intermolecular reaction of alkynamides and alkenes. *Journal of the American Chemical Society* **129**, 2230-2231 (2007).
84 Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *J Am Chem Soc* **130**, 15611-15626, doi:10.1021/ja805649f (2008).
85 Dunin-Horkawicz, S. *et al.* MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* **34**, D145-149, doi:34/suppl_1/D145 [pii]10.1093/nar/gkj084 (2006).
86 Wei, C. M. & Moss, B. Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci U S A* **72**, 318-322 (1975).
87 Furuichi, Y. & Miura, K. A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253**, 374-375 (1975).
88 Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463, doi:10.1038/nrm1403 (2004).
89 Chen, K. & Rajewsky, N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* **8**, 93-103, doi:nrg1990 [pii]10.1038/nrg1990 (2007).
90 Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nat Rev Genet* **6**, 24-35, doi:nrg1500 [pii]10.1038/nrg1500 (2005).
91 Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21**, 2300-2311, doi:21/18/2300 [pii]10.1101/gad.1564307 (2007).
92 Patel, S. B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**, 6482-6493, doi:gkn658 (2008).
93 Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186, doi:nrmicro1793 (2008).
94 Storz, G., Altuvia, S. & Wassarman, K. M. An abundance of RNA regulators. *Annu Rev Biochem* **74**, 199-217, doi:10.1146/annurev.biochem.74.082803.133136 (2005).
95 Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic Acids Res* **37**, D122-126, doi:gkn617 (2009).
96 Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29, doi:15/suppl_1/R17 [pii]10.1093/hmg/ddl046 (2006).

97 Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci U S A* **96**, 5470-5475 (1999).
98 Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387-390, doi:10.1038/35053176 (2001).
99 Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J Mol Evol* **46**, 18-36 (1998).
100 Visser, C. M. & Kellogg, R. M. Bioorganic chemistry and the origin of life. *J Mol Evol* **11**, 163-168 (1978).
101 White, H. B., 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101-104 (1976).
102 Calderone, C. T. & Liu, D. R. Nucleic-acid-templated synthesis as a model system for ancient translation. *Curr Opin Chem Biol* **8**, 645-653, doi:S1367-5931(04)00127-9 [pii]10.1016/j.cbpa.2004.09.003 (2004).
103 Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601-1605, doi:10.1126/science.1102629 [pii] (2004).
104 Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew Chem Int Ed Engl* **43**, 4848-4870, doi:10.1002/anie.200400656 (2004).
105 Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A* **106**, 7768-7773, doi:0900528106 [pii]10.1073/pnas.0900528106 (2009).
106 Crain, P. F. & James, A. M. in *Methods in Enzymology* Vol. Volume 193 782-790 (Academic Press, 1990).
107 Crain, P. F. Mass spectrometric techniques in nucleic acid research. *Mass Spectrometry Reviews* **9**, 505-554 (1990).
108 Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins* **32**, 414-424, doi:10.1002/(SICI)1097-0134(19980901)32:4<414::AID-PROT2>3.0.CO;2-G [pii] (1998).
109 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787, doi:10.1021/ac051437y (2006).
110 Woese, C. R. Bacterial evolution. *Microbiol Rev* **51**, 221-271 (1987).
111 Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* **47**, 621-669 (1983).
112 Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-4579 (1990).
113 Edmonds, C. G. *et al.* Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). *J Bacteriol* **173**, 3138-3148 (1991).
114 Ng, W. V. *et al.* Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci U S A* **97**, 12176-12181, doi:10.1073/pnas190337797 [pii] (2000).
115 Deamer, D. W. *Light transducing membranes, structure, function, and evolution*. (Academic Press, 1978).

116 I.E.D, D. in *Advances in Microbial Physiology* Vol. Volume 15 eds A. H. Rose & D. W. Tempest) 85-120 (Academic Press, 1977).
117 Soppa, J. From genomes to function: haloarchaea as model organisms. *Microbiology* **152**, 585-590, doi:152/3/585 [pii]10.1099/mic.0.28504-0 (2006).
118 Wang, G., Kennedy, S. P., Fasiludeen, S., Rensing, C. & DasSarma, S. Arsenic resistance in Halobacterium sp. strain NRC-1 examined by using an improved gene knockout system. *J Bacteriol* **186**, 3187-3194 (2004).
119 Peck, R. F., DasSarma, S. & Krebs, M. P. Homologous gene knockout in the archaeon Halobacterium salinarum with ura3 as a counterselectable marker. *Mol Microbiol* **35**, 667-676, doi:1739 [pii] (2000).
120 Zaigler, A., Schuster, S. C. & Soppa, J. Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon Haloferax volcanii. *Mol Microbiol* **48**, 1089-1105, doi:3497 [pii] (2003).
121 Muller, J. A. & DasSarma, S. Genomic analysis of anaerobic respiration in the archaeon Halobacterium sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *J Bacteriol* **187**, 1659-1667, doi:187/5/1659 [pii]10.1128/JB.187.5.1659-1667.2005 (2005).
122 Klein, C. *et al.* The membrane proteome of Halobacterium salinarum. *Proteomics* **5**, 180-197, doi:10.1002/pmic.200400943 (2005).
123 Robb, F. T., Place, A. R., DasSarma, S. & Fleischmann, E. M. *Archaea: a laboratory manual. Halophiles*. (Cold Spring Harbor Laboratory, 1995).
124 Wood, V. *et al.* The genome sequence of Schizosaccharomyces pombe. *Nature* **415**, 871-880 (2002).
125 Christie, K. R. *et al.* Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res* **32**, D311-314, doi:10.1093/nar/gkh03332/suppl_1/D311 [pii] (2004).
126 Mewes, H. W. *et al.* Overview of the yeast genome. *Nature* **387**, 7-65, doi:10.1038/42755 (1997).
127 Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563-547 (1996).
128 Bokar, J. in *Fine-Tuning of RNA Functions by Modification and Editing* 141-177 (2005).
129 Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**, 1558-1564, doi:nbt1268 (2006).
130 Pollak, N., Dolle, C. & Ziegler, M. The power to reduce: pyridine nucleotides--small molecules with a multitude of functions. *Biochem J* **402**, 205-218, doi:BJ20061638 [pii]10.1042/BJ20061638 (2007).
131 Roberts, G. C., Dennis, E. A., Meadows, D. H., Cohen, J. S. & Jardetzky, O. The mechanism of action of ribonuclease. *Proc Natl Acad Sci U S A* **62**, 1151-1158 (1969).
132 Usher, D. A. On the mechanism of ribonuclease action. *Proc Natl Acad Sci U S A* **62**, 661-667 (1969).
133 Usher, D. A., Richardson, D. I., Jr. & Eckstein, F. Absolute stereochemistry of the second step of ribonuclease action. *Nature* **228**, 663-665 (1970).
134 Grosjean, H. B. R. *Modification and editing of RNA*. (ASM Press, 1998).
135 Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res* **31**, e8 (2003).

136 Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J Bacteriol* **158**, 769-776 (1984).
137 Moore, S. D. & Sauer, R. T. Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli. *Mol Microbiol* **58**, 456-466, doi:MMI4832 [pii] 10.1111/j.1365-2958.2005.04832.x (2005).
138 Frick, D. N. & Richardson, C. C. DNA primases. *Annu Rev Biochem* **70**, 39-80, doi:70/1/39 [pii] 10.1146/annurev.biochem.70.1.39 (2001).
139 Ogawa, T., Hirose, S., Okazaki, T. & Okazaki, R. Mechanism of DNA chain growth XVI. Analyses of RNA-linked DNA pieces in Escherichia coli with polynucleotide kinase. *J Mol Biol* **112**, 121-140 (1977).
140 Hopwood, D. A. *Genetic manipulation of Streptomyces: a laboratory manual*. (John Innes Foundation, 1985).
141 Neidhardt, F. C. I. J. L. S. M. *Physiology of the bacterial cell : a molecular approach*. (Sinauer Associates, 1990).
142 Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. *Proc Natl Acad Sci U S A* **75**, 4801-4805 (1978).
143 Brosius, J., Dull, T. J. & Noller, H. F. Complete nucleotide sequence of a 23S ribosomal RNA gene from Escherichia coli. *Proc Natl Acad Sci U S A* **77**, 201-204 (1980).
144 Bayer, E. A. & Wilchek, M. Application of avidin--biotin technology to affinity-based separations. *Journal of Chromatography A* **510**, 3-11 (1990).
145 Billington, R. A., Tron, G. C., Reichenbach, S., Sorba, G. & Genazzani, A. A. Role of the nicotinic acid group in NAADP receptor selectivity. *Cell Calcium* **37**, 81-86 (2005).
146 Li, Y. & Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2,Äò-Hydroxyl Group. *Journal of the American Chemical Society* **121**, 5364-5372, doi:10.1021/ja990592p (1999).
147 Dirksen, A. & Dawson, P. E. Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconjugate Chemistry* **19**, 2543-2548, doi:10.1021/bc800310p (2008).
148 Ellen†, M. S. & Carolyn†, R. B. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie International Edition* **48**, 6974-6998 (2009).
149 Lim, R. K. V. & Lin, Q. Bioorthogonal chemistry: recent progress and future directions. *Chemical Communications* **46**, 1589-1600.
150 Liu, Q. *et al.* Structural Basis for Enzymatic Evolution from a Dedicated ADP-ribosyl Cyclase to a Multifunctional NAD Hydrolase. *Journal of Biological Chemistry* **284**, 27637-27645 (2009).
151 Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences* **105**, 15779-15784 (2008).
152 Vu, H., Stanislav†, I. P., Celia, M. & Finn, M. G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation13. *Angewandte Chemie International Edition* **48**, 9879-9883 (2009).
153 Wang, Q. *et al.* Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *Journal of the American Chemical Society* **125**, 3192-3193, doi:10.1021/ja021381e (2003).

154 Gupta, S. S. *et al.* Accelerated Bioorthogonal Conjugation: A Practical Method for the Ligation of Diverse Functional Molecules to a Polyvalent Virus Scaffold. *Bioconjugate Chemistry* **16**, 1572-1579, doi:10.1021/bc0501471 (2005).
155 Paredes, E. & Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **12**, 125-131, doi:10.1002/cbic.201000466 (2011).
156 Fromont-Racine, M., Bertrand, E., Pictet, R. & Grange, T. A highly sensitive method for mapping the 5,Ä≤ termini of mRNAs. *Nucleic Acids Research* **21**, 1683-1684, doi:10.1093/nar/21.7.1683 (1993).
157 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562-578 (2012).
158 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
159 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech* **28**, 511-515, doi:http://www.nature.com/nbt/journal/v28/n5/abs/nbt.1621.html#supplementary-information (2010).
160 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows,ÄìWheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
161 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
162 Cai, X., Hu, H. & Li, X. A new measurement of sequence conservation. *BMC Genomics* **10**, 623 (2009).
163 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415, doi:10.1093/nar/gkg595 (2003).
164 Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55, doi:http://www.nature.com/nature/journal/v452/n7183/suppinfo/nature06684_S1.html (2008).
165 Nussinov, R. & Jacobson, A. B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proceedings of the National Academy of Sciences* **77**, 6309-6313 (1980).
166 Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133-148, doi:10.1093/nar/9.1.133 (1981).
167 Rizzi, M. & Schindelin, H. Structural biology of enzymes involved in NAD and molybdenum cofactor biosynthesis. *Current Opinion in Structural Biology* **12**, 709-720 (2002).
168 Foster, J. W. & Moat, A. G. Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiological Reviews* **44**, 83-105 (1980).
169 Yokogawa, T., Kitamura, Y., Nakamura, D., Ohno, S. & Nishikawa, K. Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. *Nucleic Acids Research* **38**, e89, doi:10.1093/nar/gkp1182 (2010).

Chapter Four:

The Cellular Roles of NAD-Linked RNA and Future Directions

Ye Grace Chen, Aaron Leconte, David R Liu

Aaron Leconte isolated total RNA from *B. subtilis*, *E. aerogenes*, *V. fischeri*, *S. pombe* and *S. cerevisiae*. Ye Grace Chen conducted and analyzed all of the other experiments described.

4.1 Introduction

The development of a screen to find cellular small molecule-RNA conjugates has provided a platform for the discovery of novel RNA chemical modifications and biological roles. Further investigations into the cellular functions of NAD-linked RNA may add to the diversity of functional RNA. In this chapter, we describe the current efforts toward characterizing the biological scope of NAD-RNAs and propose future directions for the continued expansion of novel RNA modifications and functional roles.

4.2 Screen for NAD-Linked tRNAs

Previous experiments suggest that the abundance of NAD-linked RNA is about 3,000 copies per *E. coli* cell, which is comparable to the copy number of charged tRNAs. Additionally, NAD-linked RNAs are between 30 and 120 nucleotides in length, as suggested by size fractionation using HPLC. Taken together, these results suggest that tRNAs may contain the NAD modification. To test this hypothesis, biotinylated antisense probes were synthesized to individual tRNAs and incubated with cellular RNA. After streptavidin pull-down, nuclease digestion and LC/MS analysis, NAD signal was not observed in the elution fraction. Rather, the modification was only detected in the unbound and flow-through fraction. Therefore, the individual tRNAs did not contain a 5'-linked NAD.

We hypothesized that the NAD modification may be present on several or many different tRNAs, and the NAD signal from individual tRNA pull-down may be below detection limits. To address this concern, all of the biotinylated oligonucleotides that are antisense to the tRNAs were pooled for the pull-down. Again, NAD was not detected in the pull-down material and solely in

the flow-through. The lack of detection of the NAD modification suggests that the modification is present on other RNAs that are in the 30 to 120 nucleotides size range that are not tRNAs.

4.3 Detection of NAD-Linked RNA in Species Other Than *E. coli* and *S. venezuelae*

Highly conserved sequences, structures and motifs are thought to have functional value since conservation suggests there was an evolutionary advantage to maintaining them. Mutations in a highly conserved region may lead to a non-viable form or a form that is eliminated through natural selection. Alternatively, conservation may also be a result of convergent evolution, where there was an advantage to the cell for having the specific sequence, structure or motif over other versions. If NAD-linked RNA is conserved across different species and domains of life, the small molecule-RNA conjugate could be a "molecule fossil" and of an ancient origin. There may be important functional roles for the NAD-linked RNA if the conjugate is highly conserved.

There could be conservation of the presence of NAD-linked RNA, the RNA sequence, and the RNA structure. To determine whether NAD-linked RNA is present in species other than *E. coli* and *S. venezuelae*, we applied our nuclease-based screen to RNA from a variety of species. Total RNA was isolated from *B. subtilis*, *E. aerogenes*, *V. fischeri*, *S. pombe* and *S. cerevisiae*, digested with nuclease and analyzed by LC/MS. *E. coli* and bovine tRNA was also obtained from commercial sources. Our methods revealed NAD-linked RNA in *B. subtilis*, *V. fisheri* and *E. aerogenes*, as well as in bovine tRNA.

The existence of NAD-linked RNA in species spanning the bacteria kingdom and in mammalian tRNA suggests that the functional role of the small molecule-RNA conjugate may be common to many species. We hypothesize that the functional role may be fundamental to

different organisms and a greater understanding of the phylogeny of NAD-linked RNA could inform our comprehension of its roles.

Understanding sequence conversation is important for the study of sequence evolution and for the identification of functional regions of the genome.[162] To compare RNA sequences containing a 5'-linked NAD, total RNA would be extracted from the species that have been identified as possessing NAD-linked RNA. After enriching for the small molecule-RNA conjugate from total RNA, the RNA would be subjected to the adapter ligation scheme as described in Chapter 3.7. The sample could be sequenced and comparative analysis would be conducted of NAD-linked RNA from different species.

Sequence similarities serve as evidence for structural and functional conservation, as well as of the evolutionary relationships between the sequences. Consequently, comparative analysis is the primary means by which functional elements are identified. If portions of the NAD-linked RNA were identified to show high sequence conservation, those areas may be particularly important and necessary for the proper function of the small molecule-RNA conjugate.

The secondary and tertiary structure of RNA has been shown to greatly influence RNA function. Therefore, knowing the RNA structure will provide crucial insights into the way in which RNA works. Structural analysis could be performed on NAD-linked RNA to determine if there are conserved motifs between different species even if there is not sequence homology. While producing RNA high-resolution structures by X-ray crystallography and NMR spectroscopy is one method to obtaining RNA structures, this process is often slow and laborious. Several programs have been developed to analyze RNA sequences and predict the secondary and tertiary structures.[163,164]

mfold is a common tool used for secondary structure prediction of single-stranded nucleic acids. A secondary structure describes the stems of RNA, building blocks that form when two complementary regions of the sequence base pair and adopt a double-helix structure. Secondary structure can be derived from a sequence by using a combination of free-energy minimization and covariation analysis.[163] Efficient algorithms for RNA secondary structure prediction using dynamic programming methods borrowed from sequence alignment can be applied to the NAD-linked RNAs from different species.[165,166]

Secondary structures of RNA provide enough structural constraint to model tertiary structures.[164] Algorithms have been developed based on MC-fold and MC-Sym and a first-order object to represent nucleotide relationships in structured RNAs: the nucleotide cyclic motif (NCM). Adjacent NCMs share common base pairs, a property providing enough base-pairing context information to derive an effective scoring function and making possible the use of the same algorithm for predicting secondary and tertiary structures.[164] This program has been successful at predicting structures from a single sequence for fragments of up to approximately 150 nucleotides. Since NAD-linked RNAs has previously been shown to be less than 200 nucleotides, and in fact are in the 30 to 120 nucleotides range in *E. coli*, the tertiary structure prediction algorithm seems optimal for our purposes. Therefore, we can subject the sequences of NAD-linked RNAs to this algorithm in order to predict tertiary structures.

Highly conserved RNAs are often required for basic cellular function. Understanding the phylogenetic distribution of NAD-linked RNA in terms of its presence, sequence and structure will greatly contribute to our knowledge of the small molecule-RNA function.

4.4 Search for Biosynthetic Pathways Through Which NAD-Linked RNA Are Synthesized

While all known nucleoside modifications with the exception of queosine are synthesized following transcription, the biosynthetic pathways for CoA and NAD-linked RNA have not yet been elucidated. The small molecule that is found on the 5'-end of RNA may be produced through the same process as the unlinked cofactor, or there may be a parallel route for synthesizing RNA-linked small molecules.

To determine if the NAD modification on RNA is synthesized through the same pathway as the small molecule NAD, we envisioned obtaining a strain requiring nicotinamide for growth and then adding isotopically-labeled nicotinamide into the media of the mutant. If the NAD modification on RNA is synthesized through the same biosynthetic pathway as the small molecule NAD, the isotope label on the NAD from RNA should be detectable after nuclease digestion and LC/MS analysis.

In *E. coli*, the small molecule NAD can be synthesized through two different pathways, beginning with nicotinamide for the salvage pathway or aspartic acid for the *de novo* pathway (Fig. 4.1). The *E. coli* strain with mutant quinolinic acid phosphoribosyl transferase (QAPRT) can only synthesize NAD beginning with nicotinamide or nicotinic acid since it cannot convert quinolinic acid into nicotinic acid mononucleotide. QAPRT catalyzes the Mg-dependent transfer of the phosphoribosyl moiety from 5-phosphoribosyl-1-pyrophosphate to quinolinic acid, yielding nicotinic acid mononucleotide, pyrophosphate and CO_2. This reaction is the third step in *de novo* NAD biosynthesis in all organisms.[167]

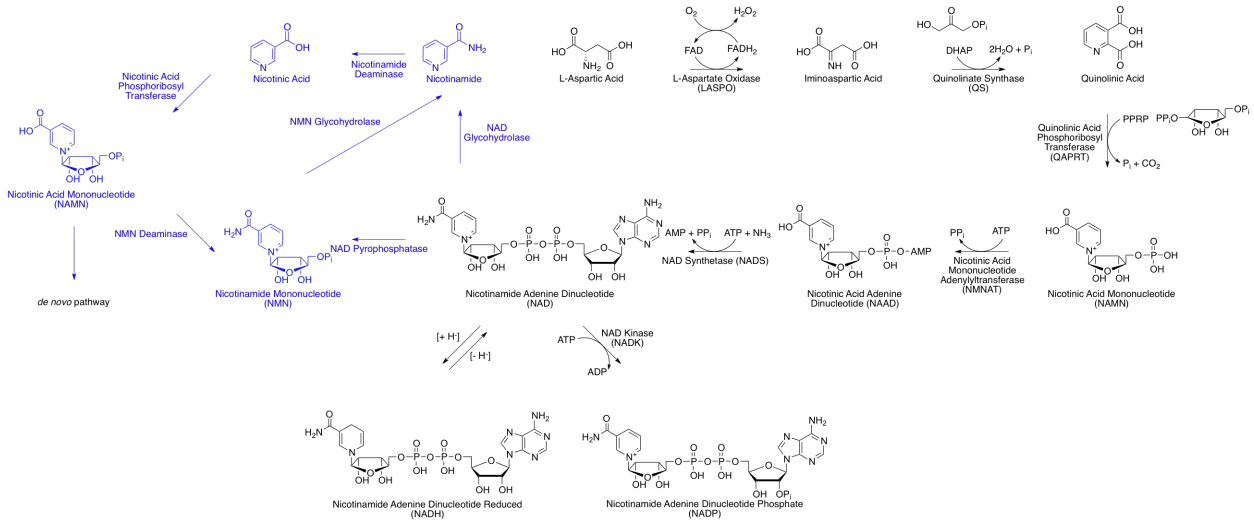


Figure 4.1. Biosynthetic pathway of small molecule NAD in *E. coli*.

We introduced isotopically-labeled (d_4) nicotinic acid into the media of the QAPRT mutant (nadC).[168] After isolating total RNA, digesting with nuclease and performing LC/MS analysis, a peak with one Da less than the expected *m/z* was observed. The unlabeled NAD was no longer detected. In parallel, the *E. coli* mutant was also cultured in media containing h_4-nicotinic acid, total RNA was isolated and digested with nuclease. LC/MS analysis revealed the presence of unlabeled NAD. We hypothesize that the labeled nicotinic acid was incorporated into NAD, and then a cycle of reduction and oxidation occurred to replace one of the deuteriums with hydrogen on the nicotinamide ring (Fig. 4.2). This would be consistent with the NAD modification being synthesized through the same pathway as the small molecule NAD, prior to installation onto RNA.

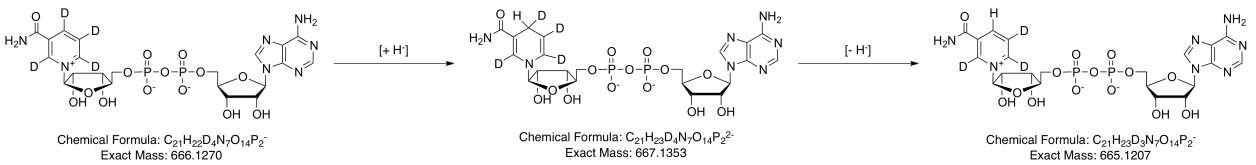


Figure 4.2. Hypothesized redox cycle of isotopically-labeled NAD prior to installation onto RNA.

4.5 Search for NAD-Binding Proteins and RNA-Binding Proteins That Are Involved in the NAD-Linked RNA Pathway

Results from the previous experiment to probe the biosynthesis of NAD from the small molecule-RNA conjugate indicate that it goes through the same route as the cofactor synthetic pathway. However, the method of installation of NAD onto RNA is not yet known and we hypothesize that there is at least one enzyme whose activity is to incorporate the small molecule onto the RNA. To identify proteins that are involved with modulating levels of NAD-linked RNA, a screen based on culturing different knockout strains can be conducted. This screen will uncover proteins that affect the synthesis or regulation of NAD-linked RNA.

Different knockout strains of proteins that are known to bind NAD or RNA were obtained. The individual strains were cultured, total RNA was isolated and digested with nuclease and LC/MS analysis was performed to measure levels of NAD modification on RNA. If the knockout mutant is of a protein that participates in either the synthesis or installation of the NAD modification, a decrease or disappearance of the NAD signal on the LC/MS should be observed. Otherwise, the NAD signal should remain constant. The extracted ion chromatogram signal of NAD was compared with known modifications to account for subtle differences in total RNA injected and LC/MS analysis variations.

Based on gene ontology databases, there are 78 proteins that are annotated as binding NAD and 190 proteins that are annotated as binding RNA in *E. coli*. None of these proteins have been identified as binding both NAD and RNA. Of the 42 knockout strains we have cultured, eight of them have shown a decrease in the NAD signal when compared with total RNA isolated from wild type *E. coli* (Fig. 4.3).

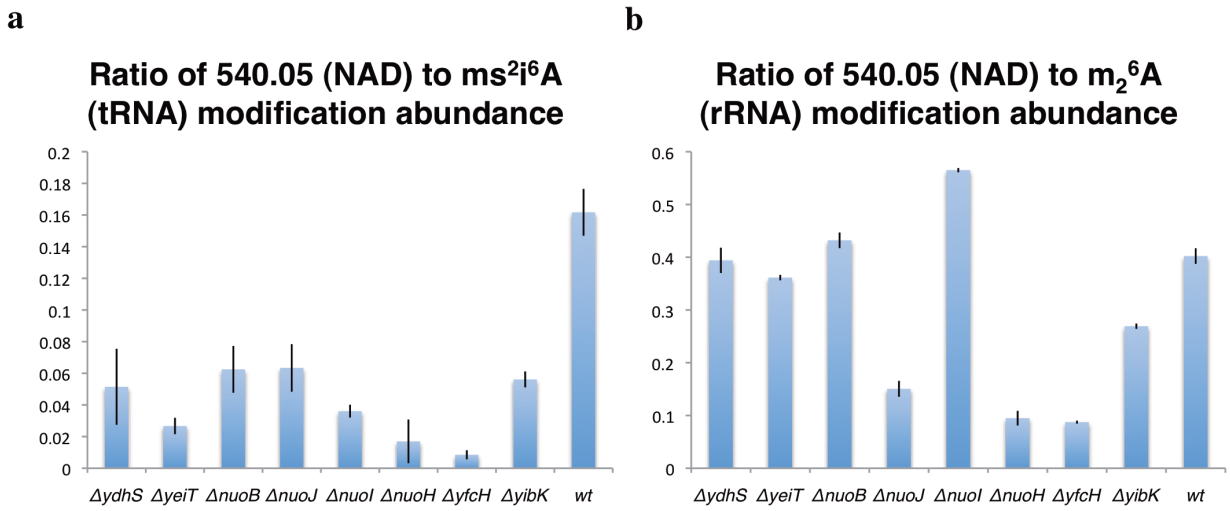


Figure 4.3. NAD signal from knockout strains of NAD-binding and RNA-binding proteins are decreased in eight strains. The ratios of NAD to two known nucleotide modifications were compared: (**a**) ms^2i^6A, a tRNA modification and (**b**) $m_2{}^6A$, a rRNA modification, in wild-type and mutant strains.

Seven of the eight strains use NAD or NADH as a cofactor for its activity, while one is a predicted rRNA modification enzyme (Fig. 4.4). To verify that these proteins are indeed involved in regulating the level of NAD-linked RNA, complementation assays are ongoing. The gene on a plasmid would be introduced into a background of the knockout strain and cultured. After isolating total RNA, digesting with nuclease and analyzing by LC/MS, we will look for rescue of the NAD levels. If the signal from NAD-linked RNA is back to wild-type, the protein that was knocked out is likely involved in processing of NAD-RNA. Further investigation into the collection of proteins that are identified as regulating NAD-linked RNA can lead to a model for the synthesis and function of the small molecule-RNA conjugate.

Strain	Annotated Function of Knocked-Out Protein
JW1658	conserved protein with FAD/NAD(P)-binding domain
JW2133	NADH-dependent dihydropyrimidine dehydrogenase subunit
JW2275	NADH:ubiquinone oxidoreductase, membrane subunit J
JW2276	NADH:ubiquinone oxidoreductase, chain I
JW2277	NADH:ubiquinone oxidoreductase, membrane subunit H
JW2301	conserved protein with NAD(P)-binding Rossmann-fold domain
JW3581	predicted rRNA methylase
JW5875	NADH:ubiquinone oxidoreductase, chain B

Figure 4.4. Annotated functions of knockout strains

NAD-linked RNA levels may be affected by environmental factors, such as temperature, pH, nutrients and oxygen levels. *E. coli* can be cultured under specific conditions and total RNA can be isolated, digested with nuclease and analyzed by LC/MS. The levels of NAD found can yield insight into its biological function since there may be dramatic responses to certain conditions. The response to environmental factors suggests that NAD-linked RNA may be involved in those pathways.

4.6 Conclusions

Given the abundance of the NAD-linked RNA conjugates coupled with the size fractionation data, we hypothesized that NAD may be present on individual or a group of tRNAs. Biotinylated antisense probes were incubated with cellular RNA and pull-downs were performed. Following nuclease digestion, the fractions from the individual pull-downs were analyzed by LC/MS. The presence of NAD-linked RNA was not detected in the streptavidin elution fraction. However, the expected NAD signal was found in the flow-through fraction, suggesting that NAD is not on the 5' end of individual tRNA. The antisense probes were also combined and used for pull-downs of all of the tRNAs from cellular *E. coli* RNA. Again, NAD was only detected in the flow-through. Taken together, the data suggests that NAD is also not

present on the group of tRNAs, but rather, is linked to another abundant RNA within the same size regime.

Other species were also cultured and total RNA was extracted to investigate the prevalence of the presence of NAD-linked RNA. The small molecule modification was detected in species spanning both the prokaryotic and eukaryotic domains of life. Applications of the nuclease-based screen revealed NAD-linked RNA in *B. subtilis*, *V. fisheri* and *E. aerogenes*, as well as in bovine tRNA. The conservation of the NAD modification suggests that NAD-linked RNA has a functional role since it has been evolutionarily preserved.

To investigate how the small molecule from NAD-linked RNA was synthesized, labeled nicotinic acid was introduced into the media of a strain that requires nicotinic acid for growth. After culturing, total RNA was isolated and digested with nuclease. By LC/MS analysis, the presence of a species that was one Da less than the expected *m/z* was observed. We hypothesized that after the labeled d_4-nicotinic acid was incorporated into NAD, a cycle of reduction and oxidation occurred to replace one of the deuteriums with hydrogen on the nicotinamide. Therefore, the NAD modification is likely synthesized through the same pathway as the small molecule NAD, prior to installation onto RNA.

We are also interested in determining the proteins that are involved with NAD-linked RNA, either through its biosynthesis, incorporation of the small molecule onto RNA or cellular function. Therefore, the knockout strains of annotated NAD-binding or RNA-binding proteins were cultured. Eight mutants that show decreased levels of NAD compared with known tRNA and rRNA modifications have been identified. We are in the process of performing complementation assays to see if these mutants can be rescued. If so, then the proteins are likely candidates that affect the levels of NAD-linked RNA.

A better understanding of the cellular roles of small molecule-RNA conjugates can add to the functional diversity of RNA. Knowing the different proteins that NAD-linked RNA interacts with, and the mechanism of interaction can help build a model for the small molecule-RNA conjugate's role. Further exploration into biological conditions that affect levels of NAD-linked RNA may inform roles of the small molecule-RNA conjugate.

4.7 Materials and Methods

General. Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification.

Nicotinamide labeling of NAD. The mutant nadC (Keio collection) was grown to an OD_{600} = 0.7-0.8 at 37 ºC in minimal media (per L: 7g K_2HPO_4, 2g KH_2PO_4, 0.5g sodium citrate $5H_2O$, 0.1g $MgSO_4$ $7H_2O$, 1g $(NH)_2SO_4$, 2.5g glucose)[168] supplemented with 10^{-6} M d_4-nicotinamide (Cambridge Isotopes, Inc.). Extraction of total RNA was as described previously.

Knockout mutant screening. Knockout strains were obtained from the Keio Collection (Yale University) and cultured to OD_{600} = 0.7-0.8 at 37 ºC. Total RNA was isolated and digested with nuclease as described previously.

RNA from other species digestion. *E. coli*, *S. venezuelae*, *B. subtilis*, *E. aerogenes*, *V. fischeri* were cultured, and total RNA was isolated and digested with nuclease P1 as described previously. Bovine tRNA (Roche Scientific Inc.) was purchased commercially and digested with nuclease P1 as described previously.

tRNA pull-downs. We synthesized antisense probes to individual tRNAs that were designed to bind to the D-loop [169] and contained a 3'-biotin. The sequences are listed below.

Ile-c: 5' TAT CAG GGG TGC GCT CTA ACC ACC T 3'

Trp-c: 5' TTG GAG ACC GGT GCT CTA CCA ATT 3'

Ala1-c: 5' CTC CTG CGT GCA AAG CAG GCG CTC TCC CA

Cys1-c: 5' AGA CGG ATT TGC AAT CCG CTA CAT AAC CGC

Asp1-c: 5' ACC CCC TGC GTG ACA GGC AGG TAT TCT AAC CG

Phe1-c: 5' AAC ACG GGG ATT TTC AAT CCC CTG CTC TAC CG

Gly1-c: 5' TCA TCA GCT TGG AAG GCT GAG GTA ATA GCC A

His1-c: 5' ACG ACA ACT GGA ATC ACA ATC CAG GGC TCT A

Ile2-c: 5' CGA CCA AGC GAT TAT GAG TCG CCT GCT CTA AC

Lys1-c: 5' CGA CCA ATT GAT TAA AAG TCA ACT GCT CTA CC

Leu1-c: 5' AGG ACA CTA ACA CCT GAA GCT AGC GCG TCT ACC AAT TC

Leu2-c: 5' GCG GCG CCA GAA CCT AAA TCT GGT GCG TCT A

Leu3-c : 5' TAC GGT TGA TTT TGA ATC AAC TGC GTC TAC

Leu4-c: 5' GGC ACT ACC ACC TCA AGG TAG CGT GTC TAC

Leu5-c: 5' CCG AGG GAT TTT AAA TCC CTT GTG TCT AC

Met1-c: 5' GAC CCC ATC ATT ATG AGT GAT GTG CTC TAA

Asn1-c: 5' CAT ACG GAT TAA CAG TCC GCC GTT CTA CCG

Pro1-c: 5' CCA CTG GTC CCA AAC CAG TTG CGC TAC CAA G

Pro2-c: 5' CCC TTC GTC CCG AAC GAA GTG CGC TAC CAG G

Pro3-c: 5' CCC CGA CAC CCC ATG ACG GTG CGC TAC CAG G

Arg1-c: 5’ TCT GCC TCC GGA GGG CAG CGC TCT ATC CAG

Arg2-c: 5’ CCC ACG ACT TAG AAG GTC GTT GCT CTA TCC

Arg4-c: 5’ CGC TCG GTT CGT AGC CGA GTA CTC TAT CCA

Arg5-c: 5’ ATT AGC CCT TAG GAG GGG CTC GTT ATA TCC

Ser1-c: 5’ CTC CGG TTT TCG AGA CCG GTC CGT TCA GCC

Ser2-c: 5’ CGC CGG TTT TCA AGA CCG GTG CCT TCA ACC

Ser4-c: 5’ CAT ACT CCC TTA GCA GGG GAG CGC CTT CAG

Ser5-c: 5’ CGT ATA CAC ACT TTC CAG GCG TGC TCC TTC

Thr1-c: 5’ CGC ACC CTT GGT AGG GGT GGG GTC CCC AGT TC

Thr2-c: 5’ CGC ACC CTT GGT AAG GGT GAG GTC GGC AGT TC

Thr3-c: 5’ GCA ACT GAC TTG TAA TCA GTA GGT CAC CAG TT

Thr4-c: 5’ GCA GCG CAT TCG TAA TGC GAA GGT CGT AGG T

Thr5-c: 5’ GCG CAC CCT TGG TAA GGG TGA GGT CCC CAG T

Val1-c: 5’ GGG AGA GCA CCT CCC TTA CAA GGA GGG GGT

Val2-c: 5’ GGT TAG AGC ACC ACC TTG ACA TGG TGG GGG T

Val3-c: 5’ GCA CCA CCT TGA CAT GGT GGG GGT CGT TGG TT

A 200 µL suspension of Streptavidin Sepharose High Performance (GE Healthcare) resin was poured into the upper cup of an Ultrafree-MC (0.22 mm; Millipore) and equilibrated with PBS. After complete removal of the buffer by centrifugation at 13,200 rpm for 1 min, 400 ml of 7.5 uM biotinylated oligoDNA solution was mixed with the resin and incubated for 10 min at room temperature. The resin was washed twice with PBS. 500 µL of total RNA was added to the resin in 2x hybridization buffer (20 mM Tris–HCl (pH 7.6), 1.8 M tetramethylammonium chloride, 0.2 mM EDTA) and incubated at 65 C for 10min. The columns were cooled to rt and

washed six times with 10 mM Tris-HCl, pH 8. The RNA was resuspended in 200 μL water, and eluted from the streptavidin by incubation at 65 C for 10min. The elution was repeated once.

4.8 References:

1 Cai, X., Hu, H. & Li, X. A new measurement of sequence conservation. *BMC Genomics* **10**, 623 (2009).

2 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415, doi:10.1093/nar/gkg595 (2003).

3 Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55, doi:http://www.nature.com/nature/journal/v452/n7183/suppinfo/nature06684_S1.html (2008).

4 Nussinov, R. & Jacobson, A. B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proceedings of the National Academy of Sciences* **77**, 6309-6313 (1980).

5 Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133-148, doi:10.1093/nar/9.1.133 (1981).

6 Rizzi, M. & Schindelin, H. Structural biology of enzymes involved in NAD and molybdenum cofactor biosynthesis. *Current Opinion in Structural Biology* **12**, 709-720 (2002).

7 Foster, J. W. & Moat, A. G. Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiological Reviews* **44**, 83-105 (1980).

8 Yokogawa, T., Kitamura, Y., Nakamura, D., Ohno, S. & Nishikawa, K. Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. *Nucleic Acids Research* **38**, e89, doi:10.1093/nar/gkp1182 (2010).